EDUCATIONAL and PSYCHOLOGICAL MEASUREMENT

A QUARTERLY JOURNAL DEVOTED TO THE DEVELOPMENT AND APPLICATION OF MEASURES OF INDIVIDUAL DIFFERENCES

VOLUME IV
1944

Copyright, 1944 by [illegible]

[illegible] INSTITUTE OF EDUCATION
LIBRARY & DOCUMENT
Unit (N.C.E.R.T.)
Acc. No. J-2.7.2.7.......
Date.......4-3-77

SMALL-PRINT EDITION

STUDENT ACTIVITIES—AN INTEGRAL PART OF STUDENT PERSONNEL WORK

[illegible] G. WILLIAMSON
University of Minnesota

Numerous writers on student personnel work have pointed out the rich opportunities for social adjustment, personality development, and citizenship training available via programs of student activities and organizations. Proponents of progressive education and teachers of social studies in elementary and high schools have developed programs geared to the objective of deriving certain skills, attitudes, and information from learning by doing in the classroom, the school, and the community. Testimony based upon observations by teachers and reports by students and alumni attest to the desirable outcomes of such participation. Studies of the persistence of activity participation of high-school student leaders in both college and the adult community indicate that social skills and participation habits, once firmly established, tend to continue as a permanent effect of training.[3] It is recognized, however, that more evaluation studies of an experimental type are needed to supplement the few now available.

With the desirability of learning social (not merely recreational) adjustment through activity participation thus indicated, it is of importance to review the methods used to enlist

---

[1] Hand, Harold C., Editor. *Campus Activities.* New York: McGraw-Hill, 1938.
[2] Lloyd-Jones, Esther McD., and Miriam, Ruth Smith. *A Student Personnel Program for Higher Education.* New York: McGraw-Hill, 1938. Chaps. VI and XI.
[illegible] Russell, John Dale, Editor. *Student Personnel Services in Colleges and Universities.* Vol. XII. Proceedings of the Institute for Administrative Officers of Higher Institutions. Chicago: University of Chicago Press, 1941. Chap. XVII.
[illegible] *Group Activities in College* [illegible]

students in such worth-while experiences. Restricting our discussion to those student activities and organizations of a non-instructional nature, we observe on most campuses marked dependence upon mass advertising appeals through word of mouth, printed bulletins, posters, and student newspapers. Notable exceptions are observed in such institutions as Antioch College, Stephens College, and George Williams College. Rather does the writer refer especially to universities in which, to a large extent, those who transfer from high school with outstanding records of leadership tend to gravitate to [illegible] opportunities while other students, inactive in high school, who need new experiences, frequently gravitate to other lines. Exceptions to this generalization are found on every campus, but the few surveys[3] made of participation [illegible] indicate that approximately one-third, varying in terms of class, and college within the University, of the students [illegible] to participate in campus affairs only by attending [illegible]. These observations are in line with the extensive surveys of interests and activities made by Lehman and Witty, which indicate that the narrowest range of interests is found at [illegible] twenty-two years. Perhaps non-participation is related [illegible] cally to some psychological factors of adolescent development which are as yet not clearly identified and understood.

To [illegible] [illegible] warrants publicity [illegible] [illegible] for leadership, or even followership in activities [illegible] little response for a number of possible reasons. A certain number prefer to maintain their home community [illegible] must devote every available hour to self-support. Some prefer to restrict participation to attending [illegible] and to cheering the football team. But anyone who [illegible] college students knows of the many students who [illegible] these subtle skills, relationships, and [illegible] as an entree to leadership and participation opportunities.

[illegible]

times the cause appears to be bizarre dress or low or high intensity of speech, and sometimes extensive shyness which may stem from myriad causes Yet many successful participants and even leaders may exhibit these same characteristics Whatever the inhibitors of desired participation, the prevalence of non-participation indicates a serious problem for many students, not only from the standpoint of mental hygiene but equally so from the point of the university's responsibility for training youth in community and citizenship responsibilities and even in the everyday living and working with others without friction. If publicity alone does not produce participation of those who desire and need such experiences, then what can be done to guide students with respect to this phase of personal development? Parenthetically, it is no small task to identify by name and person those who do not participate. Since they are not physically present at student events and affairs, special means must be employed to personalize and individualize work with them

Numerous attempts have been made in the University of Minnesota to evolve a program which would facilitate the induction of students into activities of interest to them. In these attempts there was no thought of high-pressuring all students without regard to personal preferences of some for normally restricted participation. Rather it was thought that publicity regarding the wide range of available opportunities and the increasing of this range would ipso facto attract all students who desired to participate. Our experience did not completely verify such an assumption. Even the development of a comprehensive and low-cost program of recreation in the new Coffman Memorial Student Union, while unusually successful in increasing participation in recreational affairs, has not solved the problems of an appreciable number of students who express a desire to join organizations but who do not respond to mass invitations of the publicity type.

A brief review of other program developments in this University will serve as background for the most recent innovation. In 1930 student leaders petitioned the late President Coffman to appoint a committee, with the writer as chairman, to study

in a comprehensive way the social adjustment problems of students. Prior to this time an informal group of students and staff members had constituted itself a committee to coordinate the many duplicating and competing social programs, chiefly those of a dance nature. Under the leadership of the University YMCA and YWCA a special "Fortnightly" program was developed, designed to induct the socially timid and awkward into college recreational programs. A balanced program of non-dancing games and beginner's lessons proved to be very effective until its popularity attracted those who were already well versed in social savoir faire, thus crowding out those less adept. Interesting variety in the standardized pattern of couple-dancing was introduced by the department of physical education by means of square dances, outing activities, and intra-mural sports.

The committee's survey stimulated an expansion of the social-recreational program of the Student Union and led to the appointment of a Social Coordinating Committee which continues its attempts to plan a community program for the University as a whole. This committee, working with the Calendar Committee of the All-University Student Council, has achieved some success with respect to the problem of conflicts in functions, programs and schedules of social activities.

At President Coffman's suggestion the Assistant Dean of Men developed a Leadership Course to coordinate and improve the efforts of student leaders in their conduct and leadership of organizational affairs. Little effect was noticeable with respect to increasing social participation but this group did achieve some success through its Social Etiquette or Skills Course.

The YMCA and YWCA achieved notable success in the development of small clubs and discussion groups in which social adjustment was facilitated through intimacy of contacts and concentration of program events upon specific common interests. The YWCA groups are impermanent discussion groups while the YMCA groups are fraternal but non-Greek-letter clubs of a continuing type.

The above brief review indicates some of the expanded recreational facilities Failure to participate was not caused by scarcity of opportunities But recreation is only one of the

many types of social adjustment opportunities[3] In a normal peace-time year over three hundred student organizations sponsor many types of activities in addition to the recreational program conducted by each organization largely for its own members These three hundred organizations cover the full range found on any campus, and provide rich opportunities for social adjustment, personal development and citizenship training. The intensity of appeals issued by these organizations for student workers and participants and the relatively small attendance achieved at their affairs probably indicate both the ineffectiveness of the advertising methods used and the inertia or timidity of non-participating students. Each year during Freshman Week new students are harangued and cajoled with regard to "coming out" for activities "after getting your study habits developed"

Clearly students at Minnesota are presented with opportunities and rather full information about them Should one conclude that all those who desire to participate respond to these mass appeals and that those who remain away do so through desire or are satisfactorily adjusted in social relationships? The 1932 survey of participation (to have been repeated in 1942) showed otherwise Many non-participants reported lack of knowledge of how to begin or indicated hesitancy in making the first step—perhaps mass advertising methods have reached their maximum usefulness and must be *supplemented* by other more individualized methods. Perhaps we need in college to heed the experience of group workers in community agencies with respect to the special need for personalized and individualized induction of new members into adolescent group life.[4] Perhaps we in Minnesota will find that these problems have been solved elsewhere, but an extensive search of the literature and personal observations indicates otherwise.[5] At any rate, the following developments indicate our continued search for solutions appropriate to this campus.

As of July 1, 1943, certain functions related to the supervision of student organizations and activities were regrouped administratively in a new Student Activities Bureau within the Office of the Dean of Students. The following functions were assigned to the Bureau's staff.

1 Administration of policies and regulations established by the Senate and the President concerning such matters as scholastic eligibility to participate in activities, program, time and place of recreational parties and affairs, and the like.
2 Advising with organizations regarding programs of all types, including organizations and programs designed to serve as "threshold schools" for new students and non-participating older students who need transition group experiences from high-school to college organizational participation.
3. Cooperation with faculty advisers of student organizations; development in student organizational leaders of an understanding of group and individual responsibilities for the welfare and interests of the many non-participating student followers.
4. Financial supervision of financial affairs of organizations.
5. Counseling individual students regarding the choice of and progress in extracurricular activities, organizations, and affairs.

The Bureau's staff is composed of both men and women and includes two graduate assistants This staff will deal with the organizational and activity affairs of both men and women, thereby differing markedly from the situation on many campuses where separate offices often make for difficulties with respect to affairs organized for both men and women. In such a middle western state university as Minnesota, the number of joint activities has markedly increased in the past two decades. Separate student organizations for women will, however, be continued and the special interests and problems of women will not be ignored.

The first and fourth functions listed above need no descrip tion in this paper The second function constitutes the University's equivalent of community group work in a social agency The staff members serve both as group-work supervisors and program advisers, and student leaders and faculty advisers serve as volunteer group leaders The Bureau's staff serves in advisory and supervisory capacities and not as administrators except in those cases when an organization's program is of such a nature as to jeopardize the University's public relationships. Experience indicates clearly that such instances occur infrequently when the relationships between student leaders and staff members are characterized by the term advisory as opposed to administrative. The staff serves as a resource of suggestions regarding new programs and events when student leaders and faculty advisers seek assistance. Without such new ideas, organizational programs sometimes tend to become monotonously stereotyped and fail to attract student participants. A further phase of program advising involves the coordination and community-wide planning of the programs of the several hundred organizations in order to minimize the undesirable conflicts, program and personal, which often arise as a result of failure to clear information and schedules of events.

---

[4] A similar conclusion has been reached by leaders in the field of community group work. See Joshua Lieberman, "Group Work Aims and Progressive Education," p. 67 in *New Trends in Group Work*. Ed. by Joshua Lieberman. New York: [illegible] Press, 1938. [illegible] "Group Work in Individual Guidance," p. 106 in Lieberman [illegible]

[5] [illegible] survey of extracurricular activities at nine midwestern universities [illegible]

sion. Marks concludes: "There is no university in the group which attempts to identify on a university-wide basis students who need attention with relation to their extracurricular programs" (p. 11). "In many cases it appeared that the student was not really being counseled, but was being given general advice, typically a suggestion to get into some activity group or possibly a referral to some activity director or leader. This is probably particularly true where student counselors were used who had been only superficially trained in this work. . . . Except in the case of freshman women and dormitory residents in some universities, the attempts to guide and direct students were for the most part limited to those students who sought out the help that was offered. There is reason to believe that many of the students who needed the counseling service most were not receiving it" (p. 12). See *Abstract of Ph.D. Thesis* on "The Personal Point-of-View and the Administration of Extracurricular Activities in Nine Universities" by Arlyn Marks. Urbana, Illinois: University of Illinois, 1940.

The fifth function, personal counseling, calls for some description As has been the case on most campuses, many attempts have been made at Minnesota to deal effectively with the need for assisting students to choose extra-curricular and recreational experiences in line with personal needs as outlined in preceding paragraphs. For the most part these attempts have taken the form of dispensing information about activity opportunities through the medium of bulletins, posters, "Activities Day" during Freshman Week, student counselors, and faculty counselors. Our experience has indicated that this method of counseling through activity information is probably no more effective than is the corresponding method of dispensing occupational information in the field of vocational guidance. Some students need only information about opportunities in order to choose experiences which prove to be valuable and effective. But many other students need both information and personal counseling. Just as the past decade at Minnesota

has been devoted to the developing of both clinical and faculty counseling programs in the field of educational and vocational guidance, it is expected that the present decade will be devoted to the development of counseling programs in the areas of student finances, discipline, and social adjustment through participation in extra-curricular activities Descriptions of the evolving programs concerned with financial and disciplinary counseling are in process of preparation.

While the precise dimensions to be taken and the techniques and procedures to be used in the projected activity counseling are not now known, our experience to date indicates certain things it will not become Although it must be closely coordinated with mental hygiene counseling provided by psychiatrists and psychologists and with clinical counseling regarding educational and vocational problems, activity counseling will differ from these in that it will extend beyond group therapy for serious maladjustments and beyond the measurement of aptitudes and interests. It will be as much concerned with "normal" students as with those in serious emotional difficulties. As now projected, activity counseling will provide assistance to students in the transition from the simpler society of home communities and high schools to the more complex society of adolescents which we call a university In this sense the Activities Bureau will be an agency (but not the only one) for the induction of new members into the college society For those students who are able to make their own induction transition, it will perhaps be unneeded But for the large number of students who experience difficulties in taking this step, it will offer personalized services to supplement the mass methods used at present. Such a service may be unnecessary on many campuses but it appears to be needed in this particular type of university. It should be emphasized that this type of activity counseling is supplementary to many other types of services, some of which are formally organized and others of which operate informally It should also be emphasized that participation in activities is conceived as only one of several media for achieving that degree and type of social adjustment which a particular student may desire and need There is no thought of standardizing either the means or the ends of personal devel-

opment in the area of social relationships, but rather a type of assistance which will aid the student in selecting appropriate means to achieve what he conceives to be desirable and satisfying personal objectives. Other media for achieving social adjustment will be utilized, including both organized and unorganized recreation and personal contacts, attendance at as well as participation in planning and conducting activities, residential living experiences, home relationships, and classroom and personal contacts with students and teachers

While it may appear to the reader of descriptions about this and other phases of Minnesota's personnel program that ours is a highly compartmentalized organization, such is not the case. In a large university, specialization in personnel functions is not only possible but it is also necessary from an administrative point of view if the large number of students are to be served While such specialization, with its attendant administrative departmentalization, does introduce difficulties of administrative and program coordination, discussed elsewhere,[4] the resulting improvements in effectiveness of personnel services is a worth-while gain Incidentally, these specialists supplement but do not supplant faculty counselors, who serve effectively as "general practitioners" in counseling

The Activities Bureau, as is now true of other personnel departments, serves as a focal point of a personnel function but not as an exclusive monopolizer of a function The Counseling Bureau has served as a pace setter for faculty counselors, actually helping to increase faculty participation in counseling, which now takes place literally in every building on the campus In similar manner it is expected that the new Activities Bureau will provide leadership in enlisting the faculty counselors in the use of activities as one means to further personal development of students It is hoped that through this program, students' organizations and activities will be viewed by more teachers as having worth-while educational and professional values rather than merely a distracting effect upon students' intellectual efforts.

---

[4] Williamson, E. G. "Minnesota's Program for Coordination of Decentralized Student Personnel Services." Report of the Sixteenth Annual Meeting of the American College Personnel Association, 1939, pp. 101-116.

The beginning of a program of evaluation is being set up with the establishment of the Bureau For the first year, in addition to those students who are referred or come voluntarily for counseling, a special group of several hundred students from the College of Education is being invited to use the Bureau's facilities. This group was selected because of the special motivation to be derived from the fact that effective professional adjustment of teachers is dependent to a large extent upon adequate social adjustment. Follow-up studies of these students will be made with respect to the detectable outcomes of participation Comparisons will be made between this experimental group and others composed of students who make adjustments unaided by counselors and those who fail to participate. An unusual method of identifying students who may need personalized and individualized counseling was developed as a phase of the required speech examination of all new students conducted by the Speech Clinic. The Clinic's staff is alert to the possibilities in activities for personal adjustment and personality development. Such cases are referred to the Activities Bureau for personal counseling regarding group-life adjustments. It is expected that the first few years of experience will be devoted largely to perfecting counseling procedures, delineating types of adjustment problems, and testing criteria of the outcomes of participation and non-participation. Among the criteria to be investigated are persistence of participation, degree and type of change from high school to college, subjective reports of satisfaction or dissatisfaction, alleviation of associated adjustment problems, and other factors, in relation to leadership or followership and type of activity

With the establishment of the Bureau and the initiation of an evaluation program, it is expected that the student activity phase of the University's student personnel program will gradually yield experiences indicative of its true values Not the least of these values may be the creation of awareness in counselors of the range of social adjustment problems of students and also an understanding of the therapeutic effectiveness of group adjustments as supplementary and coordinate with individualized counseling procedures and programs.

PREDICTION OF SCHOLASTIC SUCCESS IN COLLEGES OF LAW II AN INVESTIGATION OF PRE-LAW GRADES AND OTHER INDICES OF LAW SCHOOL APTITUDE[1]

WILLIAM MICHAEL ADAMS
University of Iowa

This study was undertaken with the objective of discovering additional indices of aptitude for successful law school achievement which might, in conjunction with the legal aptitude test described in an earlier paper (1), provide a more adequate basis for advising students who contemplate entering the College of Law of the University of Iowa. Specifically, it was the purpose of this investigation to determine the relative validity of pre-law grade-point averages and scores on the *Iowa Qualifying Examination* as predictive indices of success in first-year law at the University of Iowa.

Because of the variations in grading standards at different undergraduate colleges, only those students who had completed all their undergraduate work at the University of Iowa and who had complete records for one year of work in the College of Law were included in the major portion of this study These students were selected from entering classes from the years 1936 to 1939, inclusive. The number and percentages selected from each class are presented in Table 1.

The predictive indices available for the group of 152 students selected included grade-point averages for pre-law work, and scores on the *Iowa Qualifying Examination*. The *Iowa Qualifying Examination*, which had been administered to all the members of this group as entering freshmen in the liberal arts college, consisted of the *Iowa High School Content Examination*, *Iowa English Training Examination*, *Iowa Mathematics Aptitude Test*, and the *Iowa Silent Reading Test* A composite score, consisting of a weighted total of the raw scores on the four examinations, had been computed The raw scores on the individual sub-tests of the battery and the composite score had been converted to percentiles and were available in this form It was assumed that these percentiles were equivalent from year to year and, for purposes of computation, the percentiles were converted to linear scores by use of the modification of Hull's table prepared by Guilford (3).

---

[1] Second of two parts of a condensation of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, in the Graduate College of the State University of Iowa, July, 1942

TABLE 1

Number and Percentage of Students Selected from Various Freshman Classes in the College of Law

Year	Total enrolled*	Total selected	Percentage
1936	85	33	38.82
1937	96	38	39.58
1938	99	39	39.39
1939	107	42	39.25
Total	387	152	39.28

* The total number enrolled in each class included students who completed their pre-law work at other institutions in whole or in part, those who withdrew in the course of the year, and those registered as freshmen for more than one year.

On the basis of findings of similar studies made in other schools of law by Crawford and Gorham (2), Husband (5), and Jacobs (6), it was decided that it might be profitable to make separate studies of the predictive value of the total pre-law grade-point averages, the junior year pre-law grade-point averages, the pre-law grade-point averages in social sciences, and the pre-law grade-point averages in physical sciences, as well as of the scores on the *Iowa Qualifying Examination* for this group

The available criterion of success in first-year law consisted of a weighted total of the numerical grades assigned in each law course completed by the student during his freshman year in the college of law These weighted total grades were converted to ranks for each class, and the ranks were converted to linear scores by the method suggested by Hull (4) These linear scores constituted the actual criterion measures utilized in the correlational analysis

The performance of the criterion group of 152 students in the aptitude tests and pre-law areas studied is summarized in Table 2.

TABLE 2

Student Performance in the Iowa Qualifying Examination and Pre-Law Work

Predictive index	N	M	S.D	Range
Iowa Qualifying Examination.				
1 Composite score . .	152	58.97	16.32	Linear score 18–93 Percentile 5–99
2. *High School Content* .	152	59.98	16.79	Linear score 21–93 Percentile 7–99
3. *Mathematics Aptitude* .	152	58.12	18.23	Linear score 7–93 Percentile 1–99
4. *English Training*	152	53.52	16.09	Linear score 14–93 Percentile 3–99
5 *Silent Reading*	152	57.02	16.30	Linear score 11–93 Percentile 2–99
Total pre-law G. P A. . . .	152	2.58	.501	1.55–3.94
Jr. yr pre-law G. P A. . .	152	2.67	.553	1.47–4.00
Soc. sci. pre-law G. P. A. . .	152	2.65	.490	1.73–3.95
Phys. sci. pre-law G P A.	152	2.47	.647	1.17–4.00
Total pre-law G. P A	361*	2.58	.496	1.52–3.94
Jr yr pre-law G. P A.	361*	2.63	.538	1.47–4.00
Total pre-law G P A. . .	83†	2.63	.497	1.64–3.75

* A supplementary study was made of 361 students.
† A second supplementary study was made of this group of 81 students.

The mean linear scores of this group on the *Iowa Qualifying Examination* and its sub-tests were definitely superior to the scores of liberal arts freshmen as a whole, but the range was very wide, varying from scores corresponding to the fifth to the ninety-ninth percentile in the composite score. The mean total pre-law grade-point average[2] of 2.58 for this group was also superior to the grade-point average of approximately 2.20 achieved by students of the College of Liberal Arts as a whole, but probably not much superior to that of liberal arts seniors. Again the range is wide, extending from a grade-point average of 1 55 to one of 3 94 From these data it would appear that the typical University of Iowa student who enters the College of Law at Iowa is definitely superior, as a freshman, on the *Iowa Qualifying Examination* and maintains a somewhat better than average record in his liberal arts work.

Inspection of Table 3 reveals that a moderate correlation of .47 was found between the composite scores of the *Iowa Qualifying Examination* and the criterion of success in first-year law for the 152 students who had received all their pre-law training at the University of Iowa A multiple correlation of only .50 was obtained between the criterion and an optimum combination of the scores on the four sub-tests of the *Qualifying Examination* for this group. Of the indices studied for this criterion group, the total pre-law grade-point average appears to constitute the most valid single index of success in first-year law, correlating .67 with the criterion. A multiple correlation of only .68 was obtained between the criterion and an optimum combination of total pre-law grade-point averages and composite scores on the *Qualifying Examination*. Apparently addition of this latter index adds little to the predictive value of total pre-law grades.

TABLE 3

Correlation of the Predictive Indices with the Criterion

Predictive index	N	r
Iowa Qualifying Examination.		
[illegible]	152	.471
[illegible]	152	.376
[illegible]	152	.338
[illegible]	152	.418
[illegible]	152	.425
Pre-law grade-point averages		
Total pre-law work	152	.668
Junior year	152	.615
Social science	152	.642
Physical science	152	.462
Total pre-law work	361*	.579
Junior year	361*	.554
Total pre-law work	83†	.452

* A supplementary study was made of these two indices for 361 students regardless of the institution in which they had taken their pre-law work.
† A supplementary study was made of this index for 83 students who had taken all their pre-law work at institutions other than the University of Iowa.

Since it was found that, over the period of the study, students entering the University of Iowa College of Law had completed all or part of their undergraduate training at 85 different institutions, supplementary studies were made of available predictive indices for two different groups The first of these studies, based on the undergraduate transcripts of 361 students, regardless of the institution in which they had taken their pre-law work, yielded a correlation of .58 between the criterion of success in first-year law and total pre-law grade-point averages and a correlation of .55 between the same criterion and junior year pre-law grade-point averages. A second supplementary study, based on the undergraduate transcripts of 83 students who had completed all their pre-law work at institutions other than the University of Iowa, yielded a correlation of only .45 between total pre-law grade-point averages and the criterion of first-year law school performance. It would appear, therefore, that total undergraduate grade-point averages do not constitute a very effective index of scholastic success in law for students who enter the College of Law of the University of Iowa from other undergraduate institutions.

---

[2] Grade-point averages are computed at the University of Iowa by considering A=4, B=3, C=2, D=1, Fd=0.

In order to evaluate the possible efficacy of requiring a Bachelor's degree as one of the criteria for admission to the College of Law of the University of Iowa, the first-year law grades of 212 students who had entered the law school prior to obtaining their Bachelor's degrees were compared with the corresponding grades for 149 students who had received their degrees before entering the law school. A critical ratio of 2 13, significant at approximately the 2 per cent level of confidence, was found between the means of the linear criterion scores for the two groups, indicating that the former group was actually superior in first-year law school performance to the group who had received their Bachelor's degrees before entering the law school. This finding is in agreement with the results of similar studies in several other law schools. A further study of these two groups revealed that the former group was slightly, although not significantly, superior to the latter group with respect to total pre-law grade-point averages. Apparently selective non-intellective factors such as motivation, chronological age, and continuity of college attendance, as well as scholastic aptitude, must be invoked to account for the superiority of the students who entered law school after only three years of pre-law preparation

Conclusions

On the basis of the findings of the studies presented in this and a preceding article (1), the following conclusions seem justified:

1. The experimental edition of the *Iowa Legal Aptitude Test* appears to show considerable promise as a predictive index of scholastic success in first-year law for populations similar to those included in this preliminary investigation, as evidenced by correlation coefficients of from .48 through .76 between total score and first-year law achievement in five colleges Further experimentation with this test, involving larger criterion groups and utilizing more rigorous statistical techniques, should eventually result in a much shorter and more valid final edition.

2 Of the predictive indices, other than the legal aptitude test, studied in the present investigation, total pre-law grade-point averages appear to constitute the best single index of success in first-year law for students entering the College of Law of the University of Iowa from the undergraduate division of this institution The validity of this index is considerably higher for students who have taken all their pre-law training at the University of Iowa than for students who have taken part of their undergraduate training elsewhere. For students who have taken all their pre-law work at other undergraduate institutions, the predictive value of pre-law grades alone is too low to be of much value for individual prediction

3 Composite and sub-test scores on the *Iowa Qualifying Examination* do not seem to be sufficiently valid indices of scholastic success in first-year law to be very useful for purposes of individual prediction at the University of Iowa. Combination of these indices with total pre-law grade-point averages by the method of multiple correlation does not materially increase the predictive efficiency of this latter index

4. In a comparative study of the relative scholastic success in law of students who entered the College of Law of the University of Iowa after only three years of undergraduate preparation and students who received their Bachelor's degrees prior to entrance, no evidence was found to indicate the desirability of requiring college graduation as a prerequisite for admission to the college of law insofar as scholastic achievement in first-year law was concerned However, it should be remembered that the requirement of a college degree might be fully justified on other bases not investigated in this study

5 A properly weighted combination of scores on the *Iowa Legal Aptitude Test* and total pre-law grade-point averages appears to constitute the best quantitative index of success in first-year law at the University of Iowa available at present, as evidenced by a multiple correlation of .77. It is possible that the legal aptitude test scores alone will yield the best prediction for students who have taken most or all of their pre-law training at institutions other than the University of Iowa.

6. Improvement of our ability to predict scholastic success in law will probably result from experimentation with optimal combinations of pre-law grade-point averages, scores on improved tests of legal aptitude and, possibly, comprehensive college achievement examinations and objective measures of non-intellective characteristics as yet not adequately subjected to quantification Of these latter, such factors as interests, personality, perseverance, health, and personal history data in general are in need of further investigation Further analysis of the intellectual functions required in legal studies and the development of more specific tests for their measurement are needed, Such studies will be facilitated by all improvements in the reliability of the criterion of success itself

REFERENCES

1. Adams, W M "Prediction of Scholastic Success in Colleges of Law. I The Experimental Edition of the Iowa Legal Aptitude Test." *Educational and Psychological Measurement*, III (1943), 291–305
2. Crawford, A. B and Gorham, T J. "The Yale Legal Aptitude Test." *Yale Law Journal*, XLIX (1940), 1237–1249.
3 Guilford, J P. *Psychometric Methods.* New York: McGraw-Hill Book Co., 1936, 249
4 Hull, C. L. *Aptitude Testing.* New York· World Book Co., 1928, 382–390.
5. Husband, R W. "Prediction of Law School Success" *Wisconsin Law Review*, XIV (1939), 285–294.
6. Jacobs, A. C. "Comments on Dean Gulliver's Paper (from a talk before the 37th Annual Meeting of the Association of American Law Schools)" *American Law School Review*, IX (1940), 665–667

A TECHNIQUE FOR EVALUATING ASSEMBLED EVIDENCE OF POTENTIAL LEADERSHIP ABILITY

G E. MANSON AND G L. FREEMAN
Northwestern University

THE rapid expansion of our armed forces and the growth of new government agencies has created a shortage of men of proved executive talent. Consequently, the personnel expert is under considerable pressure to develop methods for selecting men of *potential* leadership ability In this task, it is essential to pool such pertinent information as is available. The present report deals with methods used in selecting a group of young men with aptitudes for business leadership It is offered as one approach to a difficult problem of personnel selection

A number of large companies with a policy of training college graduates for executive positions have been faced with the problem of appropriate selection. Such companies rarely reveal their selection techniques, and consequently there is little available information upon the subject. This was the situation presented to Northwestern University's Personnel Department fifteen years ago when it had to select from high-school seniors throughout the country, ten young men who would be subsidized in a special training course for "executive work in the field of business" The group, known as *The Austin Scholars,* were to receive a stipend covering the expenses of four years of college plus a year's travel abroad As a result of this subsidy there was no dearth of applicants, and standardized procedures had to be instituted for their progressive elimination.

The method of selecting Austin Scholars evolved around four interlocking problems. These were (1) making a job analysis of the duties and qualities of business leaders, (2) securing a population sample likely to contain potential business executives and assembling information relevant to its leadership qualities, (3) judging upon the basis of assembled evidence the relative degree to which individuals possessed the desired qualities and (4) selection from the top ranks of rater judgment upon the basis of personal interview Each of these problems will be discussed in turn.

I. *Job Analysis*

In order to make an intelligent selection of men to be trained for future executive positions, it was first necessary to ascertain the special qualities found in persons who were already successful on the job Under the auspices of its director, Dr. D T. Howard, the Personnel Department of Northwestern University prepared suggestions for a functional job analysis and submitted these to recognized business leaders for corrections and recommendations. The analysis which emerged paid attention both to the *factors held in common* by executives and all men of talent and to the *differentia* of the job,—that is, the functions, methods, and abilities which distinguished it from other types of work Among the factors common to all men of talent primary weight was given to intelligence. The differentiating features were less easily defined through study of the functions of the business executives These were listed as follows:

1 The executive must secure and evaluate information
2. He is required to be a planner and work-organizer.
3. He must engage in conferences and communicate ideas effectively
4. He must assume responsibility, making decisions and standing by them
5. He must maintain effective interpersonal contacts and have a reputation for personal integrity with both his employees and associates.

The personal qualities judged most appropriate to the successful execution of the above functions were divided into five

classes: (1) *physical* traits and endowments likely to favor effective communication of ideas, (2) *mental* abilities essential to the planning and organization of work, (3) *social* talents involved in the cooperative and directive phases of executive performance, (4) *character* traits contributing to confidence in one's integrity, and (5) *cultural* accomplishments making for personal distinction, leadership, and prestige.

II. *Sampling Procedure*

With the foregoing analyses as a guide, the Personnel Department next proceeded to secure its sample of high-school seniors and to obtain information of likely value in estimating the leadership potential of the individuals under consideration. On the assumption that high-level intelligence is a necessary starting point for further screening, this was the first point of emphasis Letters were sent to over a thousand high-school principals, asking them to post a circular describing the Austin Scholarship Program and to recommend their most promising students who met the following qualifications:

> Only those students will be considered who stand in the upper tenth of their graduating classes in scholastic achievement. Applicants must be of sound health, able to talk well and write well, must be leaders among their fellows, and must be respected in their communities for their moral and personal qualities. They must not have previously attended college elsewhere, and no person need apply unless it is his intention to prepare himself for executive work in the field of business

In the year covered by this report 441 young men signified their interest by inquiring about their eligibility for Austin Scholarships Of this number 109 did not follow their inquiries with formal application and an equal number were found ineligible for reasons ranging from lateness of application to limited scholastic requirements. This left 223 young men who satisfied all technical requirements They came from 35 different states and the District of Columbia. A folio compiled for each of these applicants carried the following information:

Form A. Personal history, including photograph, place of residence, schools attended. interests and activi-

ties, outside employment, and future occupational plans
Form B. Complete record of health from family physician.
Form C. Complete scholastic record from high-school principal.
Form D Ratings on five personality attributes (leadership, originality, popularity, speaking ability, and character) secured from three persons well acquainted with the applicant. The complete rating form, referred to as Form D, is given herein
Form E. Standardized intelligence test (*Psychological Corporation, Test VI*) given by high-school principal and returned to the Personnel Bureau for scoring.
Form F. Standardized English placement test (*Columbia Research Bureau English Test*) given by high-school principal and returned to the Personnel Bureau for scoring.
Form G. Correspondence and supplementary letters

III. *Rating Assembled Evidence*

With 223 superior applicants technically eligible for the 10 Austin Scholarships, some process of further screening had to be set up so that due consideration could be given each case in the assignment of relative rank. The technique finally developed involved five steps. (1) breakdown of list of candidates into groups of thirty each, in order of completion of application; (2) further breakdown within each group of thirty, so that three raters independently ranked the leadership qualities of each candidate relative to a group of ten applicants, (3) rejection of two-thirds (approximately) of the total group, largely upon evidence from objective tests; (4) reranking of remaining candidates (in groups of ten) by three new raters, and (5) combining, for each candidate, the rankings received from the six raters and then selecting for further study by interview those candidates who composed approximately the upper

third of the composite "order of merit" list. This procedure of progressive elimination first reduced the list of 223 applicants to 80 applicants and finally to 25 applicants It should also be

FORM D

Estimating Scale of Personal Characteristics"

Please make an estimate of Mr .. . . ..
without consulting him and return this sheet to the Director of Personnel, Northwestern University, Evanston, Illinois. The man you are asked to estimate for us is an applicant for an F C. Austin Scholarship If you are his friend, as is likely to be the case, please remember that these scholarships are designed to prepare men for executive positions in business Should the man we are considering pursue a business career? Is he really adapted to it? Might he not do better in some other vocation? Give us your best judgment on his capacities The scale below is designed to enable you to express your judgment in a form as concrete as possible.

1 Some men have distinct talents for leadership They are looked up to by their fellows, are put in positions of honor and are expected to take the lead in any enterprise that may be started At the other extreme are persons who are content to be followers, and are never asked to head up any sort of enterprise. Inbetween lie persons of varying degrees of leadership ability How would you estimate this individual in comparison with his fellows?	☐ Outstanding as a leader ☐ Is very often a leader ☐ Average. ☐ Is inclined to follow rather than to lead. ☐ Distinctly a follower	Please record here instances that support your estimate, and give reasons for your judgment.
2. Some men are independent and creative in their thinking. They have "ideas of their own," they analyze and interpret, invent, write, propose ways of doing things. Others are not original in this fashion, but seek to know how things are done by others before trying anything themselves. Your estimate of the individual should be based on what he does	☐ He is unusually original in his thinking. ☐ Distinctly more creative than the average person ☐ He is as original as most people. ☐ Inclines to depend on others for his ideas. ☐ Shows no indication to do original thinking.	"
3 Some men are quite generally pleasing to their fellows. Others make an unfortunate impression on most people whom they meet. The distinctly agreeable personality is welcome in all circles, is invited out, and has a large number of acquaintances The unpopular personality is not sought after by others, and is frequently neglected if not positively shunned. Consider the individual you are estimating in terms of his popularity; in terms of the attitudes actually shown by others toward him.	☐ One of the best-liked men in his community ☐ Quite popular ☐ Average. Welcome, but not distinguished. ☐ Not so fortunate as most people in social standing. ☐ Impresses most people unfavorably	"

FORM D—(*Continued*)

*Estimating Scale of Personal Characteristics**

4 Some men are able to speak in such a way as to hold the attention and communicate ideas clearly and readily At the other extreme there are persons whose speech is slow, hesitating, and very ineffectual In-between are varying degrees of speech ability Consider the person you are estimating in comparison with others. Do people understand him quickly and readily? Do people listen to him with ease and pleasure? Or is the opposite the case?	☐ Outstanding as a speaker ☐ Above average in ability to convey ideas. ☐ Does as well as most people. ☐ Not a very good talker ☐ Distinctly inferior in his speech.	Please record here instances that support your estimate, and give reasons for your judgment.
5 Some people arouse the greatest confidence in others. They are regarded as trustworthy in any situation, and people generally have the greatest respect for their integrity The opposite extreme is the wholly untrustworthy person who is known not to be reliable, and is never depended on. Consider this man as you know him yourself and as he is known by repute, and indicate where he stands, in the matter of reliability, in comparison with his fellows.	☐ Is most highly respected and trusted. ☐ Has a good reputation for dependability ☐ Is as reliable as most people ☐ Is frequently found to be not dependable. ☐ Reputation for reliability not good.	

Signature .. . . .

* This form was prepared by Dr D T. Howard.

noted that as a result of the methods used, each candidate in the first elimination series received a man-to-man comparison with three different groups of ten candidates each, and that *those who survived to the second elimination series received* similar man-to-man comparisons with three more groups of ten candidates each. A random sampling technique was used for placing a candidate in the ten-unit groups assigned to raters. While there was some duplication, the total constitution of each group was thus different for each rater By this method each applicant was assured of a ranking by each of three raters (6 for those who survived first elimination) and was compared with a number of other candidates, but not in groups so large as to make relative ranking difficult or meaningless The device tended to prevent inequalities which might appear if the

different groups of ten each were, as groups, unequal in ability or achievement. It also made for efficient work-organization in the Personnel Department, since it enabled raters to start on the elimination problem before all applications were completed and at times which fitted into their other work-schedules That the rating method was fairly reliable is indicated by the following intercorrelations between raters In the first ranking procedures, reliability coefficients between judges were as follows· 1 and 2, $r = +.70$, 1 and 3, $r = +.69$, 2 and 3, $r = +.73$ In the second ranking procedure, reliability coefficients between judges were as follows: 4 and 5, $r = +.54$; 4 and 6, $r = +.61$; 5 and 6, $r = +.73$ When rater evaluations given candidates who survived both competitions were compared, the following intercorrelations resulted:

TABLE 1

Agreement of 6 Judges in Total Rating of Applicants Who Survived First Elimination

Rater No.	1	2	3	4	5	6
1						
2	70					
3 . .	.69	73				
4	73	.58	.64			
5	74	72	.65	.54		
6 .. .	.74	.64	.61	.61	.73	

Agreement of average rating, group A judges (1, 2, 3) and group B judges (4, 5, 6) $r = .84$

Reliability of average rating of six judges (Spearman-Brown formula) $r = .91$

The size of the individual reliability coefficients indicates that there were relatively few wide variations in the rating of applicants. Examination of the table shows the usual tendency for reliability to increase as the final score becomes an average of a greater number of ratings A reliability coefficient of .91 would indicate that the technique of rating is about as consistent as the usual objective examination

We shall next discuss the qualifications on which the candidates were compared in the above procedure. There were 14 of these qualifications and these were considered one at a time. The procedure was to examine the assembled information contained in the folder of each of the ten members of a given

group with respect to the quality under consideration. The individual judged to have the most favorable manifestation of the quality was arbitrarily given the highest rank of 10, and the applicant with the least favorable manifestation a rank of 1 The final ranking was the summation of the individual's positions on the 14 qualities listed below, after the leadership rank had been given a triple weighting and the character score a double weighting.

GUIDE TO RATERS

Consider the assembled evidence in each candidate's folder with reference to each *of the following qualities and assign him a relative rank (10 to 1) in each case.* (*Do* not consult objective test scores.)

1 *Graduating Class Rank* (Consult Form C)

It is obvious that the high man in the graduating class of 20 has not reached his position with as much competition as has the high man in a graduating class of 200. In order to take this factor into account transfer the high-school graduating class ranks into sigma positions, and rank these standard scores from 10 to 1.

2. *Course of Study* (Form C)

We know that modern high schools in large cities are apt to be better standardized than smaller schools in remote territory, and that some states have better school systems than others We may at times have knowledge concerning the standards of specific private and public schools. Furthermore, within a given school there are "solid" and "soft" courses of study The applicant whose high-school program was filled with "solids" is to be preferred to one who offers a minimum The rating here should be made upon the basis of the quality of high school attended and the quality of the program pursued.

3 *Social Background* (Forms D and G)

While we do not care whether the applicant's family is, or is not, socially and financially prominent in the community, we do ask that the applicant have had a stimulating home environment. Has he lived in an atmosphere in which ambition was aroused and intellectual interest encouraged? Has he been led to make social contacts of some range? Above all, have his family and friends shown an interest in his work and progress?

4. *Appearance* (Forms D and G)

Include in this judgment only external appearance, since physique is rated separately Look at the applicant's photograph and judge whether he has a strong face and masculine appearance. Avoid favoring the "pretty face," and look for the face of a leader rather than of a matinee idol Many items of information will be found in letters and form D ratings which concern the general impression which the applicant makes on people at first sight.

5. *Physique* (Form B)

Is the applicant well-proportioned? Is his health good? Other things being equal, an applicant of good health and sturdy physique should be favored.

6. *Attitude toward Sports* (Form A)

Since we are looking for all-round men, the applicant's attention to outdoor sports is of importance. We know that a healthy attitude toward sports is characteristic of manliness. Men who go in for individualized sports such as golf and swimming should be considered as well as those who make competitive school teams

7 *Leadership* (Forms A, D, and G)

This item is given triple weight in our total estimate of the candidate, and so should be rated very carefully Judge to what extent the applicant has been a leader rather than a follower Special attention should be given to positions of leadership which the applicant secured through election or appointment by his fellow students.

8. *Initiative* (Forms A, D)

Does the applicant do only things he is told to do or does he start things on his own account? Is he enterprising? An organizer? Does he have hobbies that

show an investigative turn of mind? Has he shown initiative in his work, his activities, or his correspondence with us? Is he putting the application through on his own account or is somebody doing it for him?

9 *Participation* (Forms A and G)

Consider the variety and extent of the applicant's extra-curricular activities. Avoid the introvert who lives a shut-in sedentary life and favor the man with an interest in the active affairs of his community If the student has been compelled to work while in school, remember this may have interfered with extra-curricular activities An applicant who has done considerable work should have this activity scored heavily in his favor

10. *Oral Expression* (Forms A, D, and G)

That applicant should receive a high rating who has the ability to express his ideas effectively Men who have been chosen to speak on various occasions should be favored Pool all the information about the applicant's ability to speak good English in estimating this quality

11, *Written Expression* (Forms A, D, and G)

Written expression is held so important that we give a special test to all applicants. The score made on that test is not considered here. Instead, base judgment on the composition exhibited in applicant's correspondence with us, on his English record in high school, and what is said of his writing in letters.

12. *Popularity* (Forms D and G)

The applicant should be a person who is generally well liked and at the same time respected and looked up to as a leader He should be liked by all groups of people with whom he has had contact and not just by students of his own age. We especially seek to avoid the person who is looked upon as odd or freakish by his companions. That student who wins the respect of all classes of people and exhibits ability to make good contacts with all groups should be marked highest.

13 *Character* (Forms D and G)

Character is a broad word and includes morals, the observance of social conventions and personal integrity. We must also consider steadiness, reliability, dependability, and, above all, the applicant's sense of social responsibility This rating is emphasized in the final judgments by multiplying the score made by 2.

14. *Color* (Forms D and G)

This item takes into account the "individualizing" aspect of the applicant. Some men have a drab and commonplace personality; some are colorful and dynamic. Is the applicant a real individual, a vivid personality? This may frequently be judged partly upon the basis of the applicant's correspondence with us and partly from statements made about him by his teachers and friends. Remember we are seeking individuality in applicants, but not to the extent of eccentricity

It will be recalled that two further items of information have not been used in the above evaluation of applicants. These are the scores on standardized tests of mental alertness (Form E) and English (Form F) Although our present approach would be to combine such objective test rankings with the total rating score, in 1932 the objective tests were used to eliminate two-thirds of the applicants who composed the original sample. That is, those applicants who failed to place within the fifth decile[1] or better on both tests were excluded from further consideration. Such a drastic procedure was dictated largely by a desire on the part of the University that, above all, the men selected should excel in scholarship. We now question if such an overweighting of the intellectual factor was necessary in view of the high academic record of the applicants in high school

---

[1] These were in terms of Austin Scholar norms, which were 117+ for test E and 180+ for test F. Comparison with the norms originally established for both these tests showed that the decile rating for individual scores ran one to two deciles higher on the Austin norms.

Reviewing the steps taken in selecting Austin Scholars, we have:

(1) Rating all qualified applicants on fourteen qualities by three (Group A) raters.

(2) Elimination of applicants who did not have a fifth decile or better standing on both the psychological and English examination

(3) Re-rating the 14 qualities of applicants who have met the standards (imposed in 2) by three new (Group B) raters.

(4) Eliminating the lower two-thirds of applicants on the Group B ratings (imposed in 3), and calling the upper third (here 25 applicants) for personal interview

IV *The Personal Interview*

It may be recalled that up to this point, the Personnel Department had only indirect contact with applicants and had made its evaluations upon the basis of assembled evidence. It was deemed advisable, therefore, to base the final elimination upon a personal interview These interviews were handled by the Director of the Personnel Department, with some assistance from business leaders in the case of certain candidates The person conducting the interview did not know the previous rank order of the applicants referred to him. Consequently, he could obtain unbiased impressions The thing most sought in the appraisal interview was evidence of mental agility, earnestness of purpose and ability to handle oneself effectively in a difficult situation. The results of the interview were now evaluated against the evidence already assembled on the 25 candidates, and the final selection of 10 was discussed in conferences between the Director of Personnel and the six raters who had participated in evaluating evidence. In terms of their relative ranking on assembled evidence (Part III), the ten selected stood 1st, 4th, 5th, 8th, 9th, 13th, 14th, 16th, 20th, and last in the group of twenty-five.

V *Evaluation of Method*

The personnel worker is naturally interested in the outcome of the men selected by our method. Unfortunately, it has not been found possible to complete fully the file of each Austin Scholar five years after graduation But the fact that a number of Austin "rejects" enrolled in Northwestern University of their own accord makes possible certain group comparisons between scholars and non-scholars The grade-point averages for 10 Austin Scholars was 5 73 and for 20 non-scholars was 4.90 This difference approaches significance (C.R. 2 1). Furthermore, a numerical check of campus activities during the four years of college gave the Austin group an average count of 12 and the non-Austin group an average count of 7 activities The difference is not statistically significant.

Such evidence of leadership ability is difficult to evaluate and a better test is a comparison of the scholar and non-scholar groups in terms of achievement after graduation. We selected for this purpose ten men from both the scholar and the non-scholar groups whose post-school records were complete In the non-scholar group we find, some time after graduation, a salesman for a meat-packing firm, a research associate for a peace organization, a minor civil service employee, an auto salesman working for his father, a salesman for a book concern, an accountant, a chemical engineer, a high-school teacher, an office assistant in a large grocery chain, and the assistant general manager of a national manufacturing organization. The scholar group contains, besides accountants for two large industrial organizations, two specialists in industrial engineering, a high-school teacher and five men who have risen to managerial authority in large corporations. Included in this most successful group of former Austin Scholars is the associate editor of a national magazine, a purchasing agent, an executive assistant, a plant superintendent, and an assistant sales manager. While quantification of such evidence is difficult, it would appear that many Austin Scholars have done well in the business world even during an economically unsettled decade. The above evidence is influenced by the training given to Austin Scholars as well as by the selection method

From a fifty per cent return of questionnaires sent to scholar and non-scholar groups five years after their graduation from the University, it appears that the five Austin Scholars who replied earn higher salaries, do more supervisory work and receive more promotions than do the five non-scholars who replied. Averages are given in Table 2, together with Fisher's *t* tests of the significance of the differences found.

TABLE 2

Comparison of Average Salary, Number of Persons Supervised and Number of Promotions Received by Scholars and Non-Scholars

	Average yearly salary	Number persons supervised	Average number promotions
Scholars . . .	5,400	16.2	5.2
Non-Scholars	3,100	8.8	3.2
Difference	2,300	7.4	2.0
t value	6.5	10.3	3.5
Significant at	1% level	1% level	1% level

All differences are significant at the 1% level.

VI *Comment*

It is recognized that the proposed technique for selection of potential leaders has received no precise validation. Nevertheless, its internal consistency might commend it to business organizations, military boards, and government agencies who have to select from a large number of applications for a given job. By using a rating form appropriate to the job, the sorting of application papers can proceed in an orderly fashion. The authors recommend especially that applications be assembled in groups of thirty as they are received and that this number be broken into three random groupings of 10 each, so that each application is evaluated by three different raters in reference to a somewhat different sampling. By considering only the top third rankings (average of three raters) from each group of thirty, a large number of applications can shortly be reduced to more workable proportions for a second comparison by a fresh group of equally competent raters If necessary, the top third can again be skimmed from the second comparison and arranged for re-ranking a third time. By such a method, the

men finally chosen will have been ranked by a number of raters relative to several groups of applicants and in reference to a fixed set of criteria

The authors hold no special claim for the specific criteria of leadership ability used in the present selection procedure. These are probably no better nor worse than the lists of qualities now in use elsewhere in the rating of leadership potential It should be pointed out that before any such list of qualities can be used with precision, it will be necessary to determine the degree to which each item discriminates between leaders of demonstrated ability and non-leaders. Until then, the relative weighting of items on a leadership rating scale must rest largely upon the opinion of the examiner The authors feel that abstract intelligence was overweighted in the selection of Austin Scholars. The fact that this type of overweighting is characteristic of most leadership selection procedures is no justification for its continuance. It seems that personnel workers might well give further study to an alternative proposition. This proposition is that since *leadership is a social quality*, it is better judged by evidence of effective social interaction than by non-social tests of individual ability.

JUDGMENTS IN COUNSELING[1]

RALPH F BERDIE
University of Minnesota

VOCATIONAL counselors in student personnel programs attempt to assist students make judgments about their own capacities, potentialities, and personalities The counselor supplies the student with two things Primarily, he gives the student information about himself and his relationship to his surroundings The student is told how his various abilities and interests, as estimated by test scores and past achievement, compare with the abilities and interests of people in different jobs, schools, and curricula. The counselor then helps in the use of this information so that the student can make a sound judgment regarding plans for the future.

Techniques for estimating abilities and interests have been discussed by many writers and summarized by others (1) Methods of assisting students in arriving at judgments have been discussed by Williamson, Rogers, and others (4, 2) Before a student can understand his situation and plan for the future, he must make comparisons and discriminations Information that has been analyzed and evaluated results in judgments, which in turn lead to decisions, which in their turn lead to action.

An attempt to adopt methods of clinical psychology and psychiatry in vocational counseling has led to a difference of opinion regarding underlying philosophies and techniques. One group of workers emphasizes the judgments that the counselor must make. Another group expresses the opinion that it is not only unnecessary but also undesirable that the counselor make judgments. The counselor, according to the latter group, functions only to assist the student in arriving at judgments and does not provide judgments of his own which the student can accept or reject.

---

[1] This article is adapted from *Factors Associated With Vocational Interests*, by Ralph F Berdie, a Ph.D thesis on file at the University of Minnesota Library Professor D G Paterson was the author's major advisor

The view of the group deeming the counselor's judgment essential is expressed by Williamson (4, page 118):

> It should be repeated that in vocational counseling the counselor collects data about the student's potentialities and then proceeds to compare these data with the requirements of the student's expressed occupational choice In this way he arrives at a diagnosis of aptitude and a judgment as to the wisdom of that choice. *This comparison of potentialities with preferences is the important step in diagnosis.*[2] Following this step comes the interpretation to the student of the counselor's diagnosis and the cooperative planning of next steps.

The judgments studied in this investigation were judgments made by counselors about students, and the results are therefore most relevant for those counselors subscribing to the school represented by the quotation from Williamson

When a student seeks vocational counseling, he usually is considering one or more possible vocations and he is attempting to make a judgment regarding the relative appropriateness of the considered vocations before deciding to enter one. The large number of factors serving to attract a student to various vocations results in one vocation usually being more or less preferred over others. Students seldom are considering two or more vocations all of which seem of exactly equal attractiveness They have preferred vocations and they want to determine if they should enter those vocations To assist in this determination the counselor and the student organize information based upon school grades, aptitude and ability tests, achievement tests, interest tests, personality scales, hobbies, and work experience. This organization usually results in both the counselor's and the student's making judgments about the appropriateness of vocational choices The counselor's judgments may be the most effective force in molding the student's judgments.

This study sought to determine how consistent different counselors were in making judgments about vocational choices and what factors were instrumental in their arriving at these judgments. The judgments were all based upon the same material, i e, case folders, and the counselors all had gone through a series of training clinics that supposedly supplied them with a common knowledge of the relevancy of the tests used and the data collected

---

[2] Italic is the original author's.

Methods

Twenty case folders of pre-college men were selected from the files of the Testing Bureau of the University of Minnesota. The vocational choice of each of these students had been judged in the first counseling interview as either appropriate or inappropriate by the counselor. The case notes were then reviewed by a case reader who verified the judgment of the original counselor A duplicate folder was then prepared for each of the cases and all identifying data, such as name of student, name of father, etc, were removed The following material was contained in the folder in this order:

1. The preliminary interview report—containing the student's stated problem, his high-school record, and observations made by the receptionist
2. The summary profile of test scores—containing the names of the tests, the raw scores, the percentile scores, and the norm groups. Tests given at the Testing Bureau include tests of scholastic aptitude, scholastic achievement, special abilities, and personality
3. The Strong *Vocational Interest Blank* profile. Scores were recorded for twenty-eight occupational scales and two non-occupational scales, masculinity-femininity, and occupational level
4. The Kuder *Preference Record* profile, unrevised form These profiles were not available for all of the twenty cases
5. The University Testing Bureau *Individual Record Form*—a questionnaire covering family background, work experience, recreational activities, occupational plans, health history, etc.
6. A short statement of factual information obtained by the original counselor in the first interview with the student. No interpretative statements were included
7. A rating form—containing the three following statements.

   The student comes to you and says, "If you say 'yes,' I will attempt to become a(n) (*insert occupational choice*) If you say 'no,' I will not"

   In this case you are forced to make a judgment. On the basis of the contents of this folder alone, would you say:

   Yes ...... or No. . (check one)

   1. If you think someone might disagree with you on the above judgment, what do you think his basis for disagreement would be?
   2. Why do you think this basis insufficient to change your judgment?

The folders were then arranged in numerical order and given to each of the five counselors with the following instructions

> This is a necessary step in determining how judgments regarding the appropriateness of students' vocational choices are made and how consistent these judgments are. Will you please read through the twenty folders in the order in which they are given to you and after you have obtained the necessary information as you do in the interview or prior to the interview, answer the questions which are on the mimeographed sheets in each folder;
>
> It is impossible to duplicate the exact interview situation without the student, but in order to provide as much of the interview atmosphere as possible, the data observed and recorded by the original counselor have been included in the folder. Regard these observations as factual rather than interpretive, i.e, they are exact statements of description rather than what the observer thought Please be sure to answer all the questions

After each counselor had rated the twenty cases, a procedure requiring from two to four hours, the folders were checked and put in their original order New rating forms were inserted and the folders were given to the next counselor.

Results

Table 1 shows those cases judged by each counselor as having appropriate vocational choices and those having inappropriate choices The check after the number of a student indicates that the counselor judged his choice as appropriate. The absence of a check indicates that the counselor judged it as inappropriate.

Of the 100 judgments, fifty-five were in the direction of appropriateness and forty-five in the direction of inappropriateness. Of the twenty students, ten had been judged by the original counselor as having appropriate vocational choices and ten as having inappropriate choices. Eighty-one of the 100 judgments made by the five counselors agreed with the judgment previously made by the counselor who originally interviewed the student. Analysis of the data in Table 1 yields a tetrachoric coefficient of .86 between the judgments made by the original counselor and the judgments later made by the five counselors. The results indicate that in judging the appropriateness of the vocational choices of groups of students, trained counselors are able to agree.

TABLE 1

Judgments of Counselors Concerning Appropriateness of the Vocational Choices of Twenty Students

Cases judged as having appropriate choices by original counselor	Counselors who later made judgments of appropriateness				
	1	2	3	4	5
13755	x	x	x	x	
13972	x	x		x	x
14095	x	x	x	x	x
14124	x	x	x	x	x
14178	x	x	x		x
14377	x	x	x	x	x
14423	x	x	x	x	x
14443	x	x	x	x	x
14485	x	x	x	x	x
14642				x	

Cases judged as having inappropriate choices by original counselor	Counselors who later made judgments of appropriateness				
	1	2	3	4	5
13885					
13964	x	x	x		
14081	x			x	
14142		x			x
14396	x		x		
14445					
14486			x		
14600	x		x		
14640					
14769					

Of the decisions originally made in the direction of appropriate choice, 14 per cent were later judged as inappropriate. Of the decisions originally made in the direction of inappropriate choice, 24 per cent were later judged as appropriate. This suggests that counselors tend to agree more upon judgments when students' choices are appropriate than when they are inappropriate. Counselors are perhaps more willing to put their stamp of approval on the student's plans than they are to discourage him. If a doubt exists in the mind of a counselor regarding the appropriateness of a student's vocational choice, he is more apt to decide the choice warrants a try-out than to decide it does not These generalizations spring from clinical experience and the above data tend to substantiate them.

The judgments of the majority of counselors (three or more) agreed with the original classification of appropriateness-inappropriateness in eighteen of the twenty cases. The only people who considered the choice of case 14642 appropriate were the original counselor in both his first judgment and later judgment, and the case reviewer. One counselor besides the original counselor and the case reviewer considered the choice of case 13964 inappropriate, but three counselors considered it appropriate. The results do indicate that in the great majority of cases, judgments of the five counselors agreed with the decisions reached by the original counselor and the case reviewer

Fourteen of the twenty cases had originally been interviewed by counselors who later served as judges in this study. It is doubtful if the original counselors remembered the cases, as all identifying data were removed at second presentation and a period of several months had elapsed since the original interview Of these fourteen judgments, not *one* disagreed with the original decision. Even those two cases on whom the majority of counselors disagreed with the original counselor were consistently judged a second time by the original counselor. Here is evidence that a period extending from two to four months does not alter the judgments a counselor makes regarding the appropriateness of vocational choices and that counselors tend to arrive at the same judgment in the interview as they do by merely reviewing the case data and never seeing the student. These are essentially the results found by Super and Brophy (3), and it seems safe to generalize from them in the area of vocational guidance.

Up to this point, the analysis of judgments concerning appropriateness and inappropriateness of vocational choices has provided information regarding the consistency of these judgments. Little has been shown regarding the actual judgment-making processes themselves, however, and the factors considered in these judgments have not been discussed. A more complete understanding of these judgments will be obtained through a careful consideration of some of these cases themselves and the comments made by the counselors as they were making the judgments Limitations of space prevent presentation of this material for all of the twenty cases

Case number 13755 was a seventeen-year-old boy who graduated from high school one month after he came to the Testing Bureau. His occupational choice was chemical engineering, his father was dead, he had one eleven-year-old brother, and he said he would have to work about twenty-four hours a week while attending college

Test scores were as follows:

Test	*Percentile score*	*Norm group*
High-school scholarship	42	
A.C.E.	77	University freshmen
Ohio Psychological	61	U.T.B. cases*
Cooperative English	46	University freshmen
Iowa Math. Training	57	Engineering freshmen
Iowa Chemistry Training	87	Engineering freshmen
Cooperative Physics	47	National—1 year high-school physics
Cooperative Social Studies	88	S.L.A. freshmen†
Minnesota Clerical—Numbers	54	General population
Names	94	General population
Finger Dexterity	54	General population
Tweezer Dexterity	33	General population
Spatial Relations	99	General population
Revised Paper Form Board	88	Engineering freshmen
Chapman-Cook Reading Test	64	S.L.A. freshmen
Minnesota Personality Scale—		
Morale	14	S.L.A. freshmen
Social	51	S.L.A. freshmen
Family	90	S.L.A. freshmen
Emotional	60	S.L.A. freshmen
Economic Conservatism	17	S.L.A. freshmen

* University Testing Bureau cases.

† Freshmen in the College of Science, Literature and the Arts.

Strong Test scores were A on keys for chemist, farmer, carpenter, printer, mathematics and physical science teacher, forest service man; B+ on keys for dentist, engineer, Y.M.C.A. physical director, social science high-school teacher, office man; B on keys for production manager and accountant. *Masculinity-Femininity* percentile—82; *Occupational Level* percentile —2 Percentile scores on the Kuder *Preference Record* were:

Scientific, 98, Computational, 30, Musical, 8; Artistic, 86; Literary, 8; Social Service, 9, Persuasive, 44.

The following statement of factual information was obtained from data contained in the interview notes of the counselor who originally counseled the student.

> He is a husky, rather ordinary appearing young man He is a little crude in his approach but very friendly His father is dead and he lives with his mother and younger brother
>
> He will be totally self-supporting. He has saved about $100 from his paper route in the last three or four years and will continue to have a morning and evening route while going to college
>
> He says he took his first two years in high school as a joke and as compulsory education. Then he realized he must work harder but was unable to fully compensate for his previous loafing. He seems well motivated.

The original counselor and the case reviewer originally decided that the boy's choice of chemical engineering was appropriate. In making the later judgments, four of the five counselors agreed that this choice was appropriate; one decided it was not.

The dissenting counselor's response to the first question concerning the basis for disagreement was as follows:

> They would stress the adequate ability and high-school achievement factors, together with the A and B+ scores on the chemist and engineering key on the interest test. They would then suggest that the boy be given a chance to show what he could do They would also stress the good tests of background in mathematics and chemistry

In response to the second question, concerning his reasons for considering the above inadequate to change his judgment, he wrote:

> The discrepancy between H.S.R. and *A C.E.* or *Ohio* seems to indicate under-achievement of long standing. The extremely low *Occupational Level* score and the problem of complete financial support are not hopeful signs of staying power The block of skilled trades interests on the *Strong Test* is vitally important, partially because of their rarity.

Three of the other counselors mentioned the low high-school rank but did not appear to give it enough weight to have them change their judgments One counselor thought a basis for disagreement might be found in the fact that it was too early for the boy to make a choice as specific as chemical engineering. Two counselors specifically mention the boy's motivation and appear to think that the fact that his achievement tests are better than his high-school rank improves his chances in the university.

One counselor, rating the occupational choice as appropriate, recognized these negative factors

1 Average mathematics test score
2 Low high-school rank
3 Low *Occupational Level* score—"interests evidently immature"
4 C on mathematician key of *Strong Test*
5 Totally self-supporting
6 Low grades in high-school mathematics

We thus have two counselors who considered and listed the same factors, and yet one decided the choice was inappropriate, while the other wrote·

> The above is strong evidence to question his choice, but because of the positive factors which are also clearly indicated, I'd be inclined to give him the benefit of the try-out.

The above counselor's note concerning the *Occupational Level* score gives a suggestion concerning one possible source of this disagreement This counselor evidently interprets the *Occupational Level* score as an expression of interest maturity which therefore might change as the student ages, while the first counselor vaguely interprets it as an index of academic motivation

Only one of the counselors approving of the occupational choice mentioned the financial problem, and none of them emphasized it as being important The original counselor recognized the financial problem in the interview and took steps to meet it Although he realized the importance of this problem, he did not consider that the boy should stay out of engineering

This boy completed the first two quarters in the Institute of Technology with an honor point ratio of 42. His grades were·

College Algebra	D
Chemistry	C, C
Composition	D, I (incomplete)
Drawing	D, D

He dropped out of school at the end of his second quarter and has not returned.

Case number 13972 was an eighteen-year-old boy who graduated from high school six months before coming to the Testing Bureau and who was completing six months of post-graduate work in high school His occupational choice was mechanical engineering, his father was crippled with arthritis and unable to work, he had two younger sisters, sixteen and eleven years old, he had work experience drafting on an N Y.A job, and would be totally self-supporting.

Test scores were as follows·

Test	*Percentile score*	*Norm group*
High school scholarship	98	
A.C.E.	76	University freshmen
Ohio Psychological	78	U T,B cases
Cooperative English	70	University freshmen
Iowa Mathematics Training	97	Engineering freshmen
Iowa Chemistry Training	17	Engineering freshmen
Minnesota Clerical—Numbers	74	General population
Names	96	General population
Chapman-Cook Reading	20	S L.A freshman
Minnesota Personality Scale—		
Morale	10	S L.A. freshman
Social	31	S.L.A. freshman
Family	66	S.L.A. freshman
Emotional	57	S.L.A freshman
Economic Conservatism	26	S.L.A. freshman

Strong Test scores were A on keys for chemist, production manager, farmer, carpenter, printer, mathematics and physical science teacher, accountant, office man, B+ on keys for engineer, forest service man, B on keys for dentist, personnel manager; social science high-school teacher, purchasing agent. *Masculinity-Femininity* percentile—*76*, *Occupational Level* percentile—*22*.

The following statement of factual information was obtained from data contained in the interview notes of the counselor who originally counseled the student

> The boy does not have much money. His father is disabled because of arthritis. The boy has been working part-time for five months on a N Y.A job doing drafting work and making about $100 a month. He has had other odd jobs from time to time.
>
> He is intrigued with mathematics and doesn't know much about the work of an actuary He does not know that the training for this type of work is essentially through business and not through mathematics.
>
> He has applied for a scholarship open to sons and daughters of World War veterans
>
> He is socially retiring.

The original counselor and the case reviewer decided that the boy's choice of mechanical engineering was appropriate In making the later judgments, four of the five counselors agreed that this choice was appropriate, one decided it was not

The dissenting counselor wrote·

> This is a loose judgment It would actually depend upon a judgment of which field (Institute of Technology or actuarial) the boy wanted most He would be handicapped by lack of chemistry [in high school].

Three of the other counselors recognize the boy's expressed and measured interests in the business field and two of them suggest possible combinations of business and engineering. One of these two "mildly objects" that the boy hasn't quite the personality type for executive or management work. Two of the counselors indicate that the financial problem might serve as a basis for someone disagreeing with their judgment of the choice as appropriate Both say, however, they think the problem can be satisfactorily managed One counselor mentions the low *Occupational Level* scores but thinks this is overbalanced by the excellent achievement in high school This is the same counselor who in the last case interpreted the *Occupational Level* score as an index of maturity He is apparently using it here as an indicator of potential achievement! This counselor also recognizes the low score on the chemistry test but discounts it because the boy has not had chemistry in high school

In general, there are no fundamental disagreements among the counselors regarding this case They all tend to recognize that the boy has several potentialities and is a good college risk Although one counselor indicates he would not encourage the occupational choice, he later qualifies this.

This boy completed the first year in the Institute of Technology with an honor point ratio of 2.11. His grades were:

Algebra	A
Trigonometry	A
Analytic Geometry	B
Chemistry	B, B, C
Composition	A, B, B
Drawing	B, B, C

He is now making satisfactory progress in his engineering course

Case number 14095 was an eighteen-year-old boy who graduated from high school the same month he came to the Bureau. His vocational choice was business, the occupation of his father. His test scores were as follows.

Test	*Percentile score*	*Norm group*
High school scholarship	30	
A C.E	53	University freshmen
Ohio Psychological	45	U T B cases
Cooperative English	49	University freshmen
Iowa Mathematics Aptitude	64	University freshmen
Iowa Mathematics Training	8	University freshmen
Cooperative Natural Sciences	24	S.L.A freshmen
Cooperative Social Studies	87	S.L.A. freshmen
Minnesota Clerical—Numbers	94	General population
Names	89	General population
Mechanical Assembly	56	General population
Finger Dexterity	40	General population
Tweezer Dexterity	86	General population
Spatial Relations	86	General population
Chapman-Cook Reading	70	S.L.A. freshmen
Minnesota Personality Scale—		
Morale	93	S.L.A. freshmen
Social	96	S.L.A. freshmen
Family	31	S.L.A. freshmen
Emotional	84	S.L.A freshmen
Economic Conservatism	53	S.L.A freshmen

Strong Test scores were· A on keys for personnel manager, real estate salesman; B+ on keys for social science high-school teacher, accountant, office man, sales manager, life insurance salesman, B on keys for production manager, Y.M.C.A. secretary, purchasing agent *Masculinity-Femininity* percentile—*76*; *Occupational Level* percentile—*66*. Percentile score on the Kuder *Preference Record* were: Scientific, 42, Computational, 79; Musical, 10, Artistic, 7, Literary, 62, Social Service, 60; and Persuasive, 99

The following statement of factual information was obtained from data contained in the interview notes of the counselor who originally counseled the student.

> He is a young man of short stature, not too neatly dressed, has a rather informal manner in the interview He is rather ordinary in appearance. He would like to be self-supporting while going to school He has about $250 saved now and might earn enough before starting at the University to continue without working while in school. His father would help him but he dislikes asking his father for help
>
> The father and mother were divorced about three years ago. The father is the president of a business. There are no other children in the family and the boy lives with his mother and gets along well with both parents.
>
> In high school he says he studied just enough to get by but says he realizes now that that was not so smart

The original counselor and the case reviewer decided that the boy's choice of business was appropriate. In making the later judgment, all of the five counselors agreed that this choice was appropriate. Four of the five said they did not think the boy's chances of graduating from the University were good but their attitude is expressed by this counselor.

> I can't see any basis for disagreement since the job is waiting for him when he is through with school. However, I do not think he'll graduate from the University

Again in this case there is general agreement among the counselors. It is doubtful if the boy's plans would have been influenced differentially by the counselors as they were all ready to approve of his vocational choice. This case is quite clear-cut.

This boy entered the College of Science, Literature and the Arts and at the end of his first quarter had an honor point ratio of .00. It is doubtful if his academic "flop" will seriously interfere with his vocational adjustment His grades were:

Composition—cancelled	
Economics	D
Political Science	C
History	F

He did not return to the University after his first quarter.

Case number 13885 was an eighteen-year-old boy who came to the Testing Bureau the month he graduated from high school. His father was a salesman. The boy had one younger sister The family was going to finance his college education

He had no work experience outside of working for his father and his vocational choice was advertising.

The test scores were:

Test	*Percentile score*	*Norm group*
High school scholarship	30	
A.C.E	8	University freshmen
Ohio Psychological	10	U.T.B cases
Cooperative English	2	University freshmen
Iowa Mathematics Aptitude	7	University freshmen
Iowa Chemistry Aptitude	19	University freshmen
Cooperative Natural Science	62	S.L.A. freshmen
Cooperative Social Studies	28	S.L.A. freshmen
Minnesota Clerical—Numbers	63	General population
Names	42	General population
Mechanical Assembly	24	General population
Finger Dexterity	29	General population
Tweezer Dexterity	46	General population
Spatial Relations	32	General population
Manual Dexterity	13	General population
Revised Paper Form Board	15	Engineering freshmen
Chapman-Cook Reading	4	S.L.A. freshmen
Minnesota Personality Scale—		
Morale	30	S.L.A freshmen
Social	9	S.L.A. freshmen
Family	10	S.L.A freshmen
Emotional	3	S.L.A. freshmen
Economic Conservatism	9	S.L.A. freshmen

Scores on the *Strong Interest Test* were. no A ratings, B+ on keys for farmer, purchasing agent, real estate salesman; B on keys for personnel manager, office man, president of a manufacturing concern. *Masculinity-Femininity* percentile—57; *Occupational Level* percentile—40. Percentile scores on the Kuder *Preference Record* were. Scientific, 57; Computational, 45, Musical, 1, Artistic, 58; Literary, 15, Social Service, 17; Persuasive, 52

The following statement of information was obtained from the data contained in the interview notes of the counselor who originally counseled the student.

> He is a very immature boy and his mother came with him the first day He gives the impression of not knowing much about anything. He doesn't react favorably to General College.

The original counselor and the case reviewer originally decided the boy's choice of advertising was inappropriate. In making later judgments, all of the five counselors agreed that this choice was inappropriate Three of the judges could see no

reason why anyone would disagree with them. Two of the counselors suggest he might be able to handle a lower level job of advertising.

> They might think he could do a low-grade type of advertising (show card writing, window displays) which may be what he means by advertising man. He could never get the degree from the advertising sequence in the Business School because of various reasons. I doubt whether he'd be much good even at a low level of advertising in view of the poor manipulative skills and ability to work in spatial relationships.

The judges appeared to have very little difficulty in arriving at a decision on this case. He entered the General College and at the end of the first year took three comprehensive examinations, receiving a C in Social Civics, an F in Vocational Orientation, and an F in Individual Orientation. His course grades were:

Individual Orientation	C		
Current History	B	C	B
Government Studies	B	B	
United States in World Civilization	B	C	
Vocational Orientation	D		
Human Biology	C	C	
Oral Communication	C		
Psychology	C		
Physical Science	B		
Human Development	C		
Economics	C		

Case number 13964 was a seventeen-year-old boy who came to the Bureau the month he graduated from high school. His father was a car and truck mechanic and there were two younger brothers and a younger sister in the family He had no unusual work experience and his vocational choice was architecture. Test scores were as follows:

Test	*Percentile score*	*Norm group*
High school scholarship	99	
A.C.E.	53	University freshmen
Ohio Psychological	67	U.T.B. cases
Pressey Classification	84	General population
Cooperative English	23	University freshmen
Iowa Mathematics Training	91	University freshmen
Iowa Chemistry Aptitude	80	University freshmen
Meier-Seashore Art	86 (exceptional)	
Minnesota Clerical—Numbers	73	General population
Names	47	General population
Revised Paper Form Board	83	General population
Chapman-Cook Reading	13	S.L.A. freshmen
Minnesota Personality Scale—		
Morale	13	S.L.A. freshmen
Social	69	S.L.A. freshmen
Family	42	S.L.A. freshmen
Emotional	59	S.L.A. freshmen
Economic Conservatism	32	S.L.A freshmen

Scores on the *Interest Test* were. A on keys for printer, mathematics and physical science teacher; B+ on keys for farmer, personnel manager, accountant, office man, B on keys for carpenter, forest service man, Y.M.C.A. physical director, Y.M.C.A. secretary, social science high-school teacher, *Masculinity-Femininity* percentile—30; *Occupational Level* percentile—22 Percentile scores on the Kuder *Preference Record* were: Scientific, 50; Computational, 65; Musical, 63; Artistic, 90, Literary, 35, Social Service, 3; and Persuasive, 30

The following statement was obtained from the data contained in the interview notes of the counselor who originally interviewed the student.

> There is a financial problem here, to some extent. He has applied for N.Y.A. assistance He claims he enjoys art very much and has done a great deal of poster work and layout work. He has a chance for a scholarship at the Institute of Art but will not accept it.

The original counselor and the case reviewer decided the boy's choice of architecture was inappropriate. In making the later judgments, two of the counselors agreed that the choice was inappropriate, three decided it was not. All the counselors rating the choice as appropriate noticed the absence of measured interests but made the following statements

"No confidence in Strong for architecture." "Kuder test shows 'artistic'; has art skills, has necessary ability. The Strong test may be in error." "The Strong is against me but his experience (making posters) is good enough to lead me to let him try it The other interest test supports his choice."

One of the counselors rating the choice inappropriate cited in support the great concentration of skilled trades interest, the low *Occupational Level* score, the poor home and cultural background, and the financial problem. The superior high school rank was discounted in light of the relatively poor school the boy attended The other counselor rating the choice as inappropriate admitted the boy probably had enough ability to get the degree and that the choice might be appropriate from the standpoint of job opportunities He added, however.

> He probably is not outstanding enough in ability to make his living in this field, which offers limited opportunities to all but the very exceptional person In view of the lack of Strong's measured interests, the low level of occupational aspiration, I'd say he should probably get mechanical drafting training or commercial art training.

This counselor, who is now interpreting the *Occupational Level* score as a measure of occupational aspiration, is the same one who previously interpreted it as a measure of interest maturity! This boy did not enter the University

In this case, the basis for disagreement is quite clear-cut. Two of the counselors, chiefly on the basis of the interest test, would discourage the boy from entering architecture The other three would discount the lack of measured interests and not discourage the boy. A tendency has also been found for some of the counselors to consider some of the keys of the *Strong Test* valid and some not. The factors which help the counselors determine which keys are *valid* have not been identified.

Case number 14081 was a nineteen-year-old boy who came o the Bureau one year after graduating from high school He ill have to help finance his college education His vocational hoice was metallurgical engineering

The test scores were as follows

Test	*Percentile score*	*Norm group*
High school scholarship	13	
A C E.	47	University freshmen
Miller Analogies, Form B	68	Education freshmen
Ohio Psychological	21	U T.B cases
Cooperative English	3	University freshmen
Iowa Mathematics Aptitude	34	University freshmen
Iowa Mathematics Training	32	University freshmen
Iowa Chemistry Aptitude	68	University freshmen
Iowa Chemistry Training	65	University freshmen
Cooperative Social Studies	85	S L.A freshmen
Cooperative Contemporary Affairs	57	S.L.A sophomores
Minnesota Clerical—Numbers	34	General population
Names	29	General population
Chapman Cook Reading	5	S.L.A. freshmen
Minnesota Personality Scale—Morale	22	S L.A. freshmen
Social	47	S L.A. freshmen
Family	66	S.L.A freshmen
Emotional	57	S.L.A freshmen
Economic Conservatism	49	S.L.A. freshmen

Scores on the *Strong Test* were: A on keys for engineer, chemist, farmer; B+ on no keys, B on keys for architect, mathematician, printer, mathematics and physical science teacher, purchasing agent. *Masculinity-Femininity* percentile—65; *Occupational Level* percentile—66 Percentile scores on the Kuder *Preference Record* were· Scientific, 70; Computational, 91; Musical, 50, Artistic, 55; Literary, 5; Social Service, 9; Persuasive, 60.

The following statement of information was obtained from the data contained in the interview notes of the counselor who originally interviewed the student

> He lives with his mother and sister, both of whom are working. His father is dead. He applied for admission last year but was referred to the General College. He did not wish to go there so he did not come to school He has an unusually strong vocational fixation.

The original counselor and the case reviewer decided the boy's choice of metallurgical engineering was inappropriate In making the later judgments, three of the counselors agreed with this decision while two counselors judged the choice as appropriate Two of the counselors rating the choice as inappropriate said they would expect no disagreement. The other counselor rating it as inappropriate said that, in the light of the boy's apparent *strong drive for engineering, and "with the* equivocal results on the Strong's," some might claim he should be given an opportunity in engineering. This counselor does not expound the mentioned equivocalities of the *Strong Test* He decides, however, that, without an interview, the above factors cannot be given much weight. Both counselors rating his choice as appropriate recognize the boy's limited abilities and one mentions his relatively weak background in mathematics and chemistry. One of the counselors writes·

> They would question his ability. His vocational choice has changed since he left high school That might help him to work to capacity Then I think he could make it . I would improve his motivation, I hope, and if so, he'd make it He lacks verbalistic skills. This would not be such a handicap to him in the Institute of Technology

This boy did not register in the University.

In this case the basis for disagreement is again quite clear. The boy has interests congruent with his choice but abilities and a background that are not encouraging All the counselors apparently recognized these factors, but three of them decided engineering was not a satisfactory choice, while two of them decided to take a chance, not place too much weight on the ability factor, depend on the interest and motivational factors, and let the boy try it.

Summary of Analysis of Judgments

The preceding discussion has covered much territory and many inferences have been drawn. Inferences drawn from the six cases discussed and the fourteen remaining cases are not conclusions based upon experimental results but are generalizations based upon observations of only a few cases Often they *are merely clinical hypotheses that should be investigated.* Some of them are nothing more than verbalizations of conditions generally known to exist but often not explicitly recognized

1. In making judgments regarding the appropriateness of students' vocational choices, trained counselors agree with the original judgment 84 per cent of the time. The tetrachoric coefficient between judgments made by different counselors is .86.

2. The majority of five trained counselors agreed in making judgments regarding the appropriateness of students' vocational choices upon the basis of the case notes with the decisions of the counselor who originally counseled the student and the case reviewer 90 per cent of the time

3. In fourteen out of fourteen cases, the counselor who originally counseled the student and judged the appropriate*ness of his vocational choice at that time made the same judg*ment from the unidentified case folder several months later.

4 Counselors tend to approve of vocational choices oftener than they disapprove of them.

5. The *Occupational Level* score has been interpreted as a measure of interest maturity, as a measure of motivation, as

an index of occupational aspiration, and as a cause for poor achievement in high school

6 Some counselors are reluctant to have a student of high ability enter business even if other factors suggest the advisability of such a choice

7. When both measured and expressed interests are in business, counselors will seldom discourage a student from entering business. They may disagree on the level of training desired

8. If ability and interest factors indicate success in an occupation, counselors tend to think meager informational backgrounds can be compensated for.

9 The appropriateness of a student's choice cannot easily be agreed upon when the occupational requirements, duties, and status are poorly defined

10 The greater the apparent contradiction between the student's test scores, the greater will be the disagreement regarding the appropriateness of his vocational choice

11. Some counselors interpret more rigorously scores on some scales of the *Strong Test* than they do scores on other scales

12. Counselors do generalize results of the *Strong Test* to occupations *for which there are no keys Often they disagree* upon the probable pattern people in a given occupation would obtain.

13 Lack of ability and absence of measured interests in an occupation almost always lead counselors to disapprove of a student's choice of that occupation

14. If a student's aptitudes are too low, a counselor will disapprove of his vocational choice even if he has measured interests in that occupation

15. Relatively little attention was paid to personality test scores in judging the appropriateness of vocational choices. Much attention was given to tests of ability, information, interests and high-school achievement.

16. Interest tests are used in determining occupational areas; ability and information tests and high-school achievement are used in determining levels of training.

17. Counselors *make decisions regarding* the appropriate-

ness of vocational choices based upon generalizations which are often without support.

18. Sometimes when counselors consider the same evidence in making judgments on the appropriateness of vocational choices, they will not reach the same conclusion. They sometimes place different weights upon tests at different times, interpret them differently, and disagree with each other regarding these interpretations.

Evaluation of these results might cause one to question if counseling based upon such judgments can be more than systematic guess-work. In light of the statistical results, however, it must be concluded that regardless of the discrepancies in interpretations and regardless of some of the peculiar logic used by the counselors, the final outcomes at which they arrive do agree. We must remember that in the above cases the counselors were forced to make a decision. In many actual interviews, the counselor can suggest try-outs of various kinds that provide additional information in terms of actual success or failure before a decision has to be made. Often the counselor cannot make a decision and does not have to put himself on a limb as he did here. The student is also an important factor in arriving at a final decision in the actual counseling situation. The responsibility for making the decision is eventually his, and counseling which is more "non-directive" in character may not force the counselor to make a decision at all As long as the counselor is to present test data to the student, however, he will most likely have to interpret it as either supporting or not supporting the student's choice or else interpret it as being completely irrelevant.

REFERENCES

1 Bingham, W V *Aptitudes and Aptitude Testing* New York, Harper and Brothers, 1937.

2. Rogers, C. E. *Counseling and Psychotherapy* Boston Houghton-Mifflin, 1942.

3. Super, D E. and Brophy, D A. "The Role of the Interview in Vocational Diagnosis" *Occupations*, XIX (1941), 1–5.

4. Williamson, E G *How to Counsel Students* New York, McGraw-Hill, 1939

THE DEFINITION OF SOCIAL EFFECTIVENESS AND LEADERSHIP THROUGH MEASUREMENT*

HENRY C. LINK

The Psychological Corporation

ENOUGH EXPERIMENTAL work has now been done to permit a definition of social effectiveness and leadership, and to justify such conclusions as the following (1) Social effectiveness consists of certain habits and skills which can be acquired by practice, (2) Leadership is an aspect of social effectiveness—leaders are not merely born, they are persons who have developed social effectiveness to an unusual degree, (3) That "only a few are born to be leaders, the great majority must be followers," is a fallacy—most people can develop and use skills of leadership if they will The basis for these statements and a set of guides for the development of leadership and social effectiveness will be presented here.

Measurement

The scientific definition and explanation of human traits depends upon the development of tests which will measure those traits The wide variety of measuring instruments developed in recent years under the broad heading of personality tests has contributed much to our understanding of social effectiveness. These tests, or scales, or inventories, center around the concept of personality as the social-stimulus value of a person Whereas recently there has been a trend to make the term, personality, synonymous with person or individuality, measurement has developed according to the more specific concept of personality in terms of social effectiveness.

---

*This paper was prepared for PSYCHOLOGY SCIENCE AND PRACTICE edited by E. R Henry and Douglas H. Fryer, publication of which by Farrar and Rinehart has been delayed by the war Publication of this paper is with their permission.

From the many excellent personality tests, the following have been selected to illustrate certain positive and negative aspects of social effectiveness. The Moss *Social Intelligence Test* (1930), because it illustrates the difference between social knowledge and social effectiveness; The Allport *Ascendance-Submission Scale* (1928), because it illustrates a widespread belief that people fall into two broad classifications, those who dominate and those who are dominated, The *Vineland Social Maturity Scale* (Doll, 1936), because more than any other test or inventory it exemplifies the concept of social effectiveness as the acquisition of specific habits and habit patterns through an orderly process of learning, and the Link *Personality Quotient Test* (1935), because it is the only test the validity of which has yet been established in terms of leadership.

Social Intelligence.—Social intelligence is not synonymous with social effectiveness An investigation of practically every experiment in which the results of personality tests were compared with those of intelligence tests, tests of academic achievement, years of education, or records of scholastic standing, reveals the almost complete absence of significant correlation between them (Link and Roslow, unpublished), that is, measures of social behavior do not give correlations with measures of intelligence or academic achievement. On the other hand, significant inter-correlations are found among nearly all of the various tests of personality, such as extraversion-introversion, ascendance-submission, sociability, sense of inferiority, emotional stability. This broad experimental result makes inevitable the conclusion that personality tests measure a common factor quite distinct from that measured by all forms of intelligence and academic achievement tests The I.Q. (intelligence quotient) or the A.Q. (achievement quotient) measures one aspect of individuality, the S.Q. (social quotient) or the P.Q. (personality quotient) measures a quite different aspect

This difference between intelligence and knowledge on the one hand, and habits of social behavior on the other, becomes dramatically clear with reference to the *Social Intelligence Test*. This test correlates not with the personality tests but with the intelligence tests. This test emphasizes an understanding or knowledge of social behavior, whereas the emphasis in the Allport *A–S Scale*, the *Vineland Social Maturity Scale*, the Link *P.Q. Test*, is upon specific social habits. The former stresses theory and judgment, the latter practice. Social effectiveness consists of a peculiar collection of habits and skills, the common denominator of which is their practical effect on other people.

Ascendance and Submission.—The *A–S Scale* codifies a concept of social effectiveness quite widely accepted both by scientific psychologists and the public According to this concept, some people dominate others in their social relationships, while some are submissive, bashful, or retiring. This idea was popularly expressed in Scott's pioneer book on influencing men in business (1911) and is also exemplified in another early book by Strong on the psychology of selling life insurance (1922).

The ascendance-submission concept of social effectiveness, crystallized by the *A–S Scale*, is carried over into the *Bernreuter Personality Inventory* and a number of other modern personality scales The use of such scales implies that a person must either dominate others or be dominated by them. This fits in with the popular fallacy that people naturally divide into two groups, leaders and followers. To be sure, test scores show many people at the happy medium, neither ascendant nor submissive. Nevertheless, the implication is that both extremes are essential We are here confronted by two aspects of psychology, sometimes called pure and applied psychology. One is the impersonal science of people as they are, the other is the science of people as they ought to be. According to the first, many people simply are submissive and many are dominant, and these are the cold facts The second emphasizes the study of people's capabilities and tries to establish standards or ideals for these capacities. In the light of such tests and experiments, a definition of social effectiveness in terms of those who dominate and those who are dominated, those who are leaders and those who are not, is inadequate.

Social Maturity.—The *Vineland Social Maturity Scale* is unique in the extent to which it embodies the measurement of social effectiveness as a result of learning. Its construction makes this clear Practically all of its 117 items refer to cific habits and habit patterns graduated according to their complexity into 17 age periods All of these habits are of a *kind which can be or must be acquired by practice.* The more of them the individual acquires, the more socially mature he becomes, no matter what his chronological age However, the grouping of habits by ages implies that the normal person will acquire the habits characteristic at his age. There is no implication of social dominance or submission, except in so far as the person who has acquired these habits well may be superior to the person who has not learned them.

This scale is unique also in the extent to which it consists exclusively of objective habit or behavior items, at least in intent Unlike so many personality tests and scales, it asks no questions about how the individual feels, whether he is ever troubled with a sense of inferiority or whether his feelings are easily hurt. Social maturity is defined entirely in terms of overt behavior, and not in terms of ambiguous introspective responses to subjective questions Through this scale Doll almost seems to say. “If a person will take care of his habits, his feelings will take care of themselves.” If this was not Doll's intention, it certainly was that of the author of the *P Q. Test,* in which the items are almost entirely habits and activities, and where such subjective traits as emotional stability, sense of inferiority, of security, and of social acceptance were deliberately omitted

A person's feelings or emotions as a result of social situations are unquestionably important. However, both psychological theory and the trend in the development of tests and in clinical work justify the conclusion that feelings of social inadequacy leading to a general emotional instability are not the cause of social impotence. They are the result of failure to acquire adequate social skills Therefore, while the emotions of frustration may give us broad clues to the problems of social competence, the achievement of social effectiveness must be defined in terms of the habits and skills which make it possible.

Doll's *Scale* is not divided into traits, as are many personality tests. Nevertheless, some of the items in his scale include such a complex range of habits that they might almost be said

to represent a unit trait. For example, the item, “assumes personal responsibility,” may well include such a net-work of habits as would justify the trait description of self-reliance. In fact, most of the traits which other tests make explicit, plus a good many more, are implicit in the items of the *Social Maturity Scale.*

There is one important exception, and that is the collection of habits which have to do with adjustment to the opposite sex. Nowhere in this scale is the specific and distinctive character of these habits recognized. On the other hand, clear recognition is given to the importance of habits and skills which contribute to social effectiveness by enabling a person to be independent of people, to get along without them. In children this consists of being able to dress alone, to eat without help, to go places alone In adults it consists of such items as “creating own opportunities” and “systematizing own work.” The importance of such habits in terms of social effectiveness is so obvious that it is easily overlooked For instance, how effective, socially, could a healthy adult be who had never learned to dress himself? On a higher level, this fact is expressed by Emerson's epigram: “The test of true friendship is the ability to do without it.” A person who is entirely dependent on his friends, like an adult who has never learned how to get along without his mother, is limited in his social effectiveness.

The person interested primarily in the practical problem of becoming more effective socially would do well to use the *Social Maturity Scale* as a guide. From the standpoint of scientific psychology some of its details are open to criticism. Nevertheless, it represents one of the soundest attempts to measure social effectiveness thus far The next great scientific task is that of delineating the habits of social effectiveness with a curriculum and series of texts for instruction comparable to those in the academic field today Doll's *Social Maturity Scale* points the direction of this development.

The P Q.—The *Personality Quotient Test* was constructed to express the definition that: “Personality is measured by the extent to which the individual has acquired habits and skills which interest and serve other people” (3). It might just as

well have been given some such name as *Social Effectiveness Quotient Test,* since it is concerned entirely with habits and skills in this category. These habits are classified in four groups in the test, representing the following four traits: (1) habits of social initiative, (2) habits of self-determination, (3) habits of economic self-reliance, (4) habits of adjustment to the opposite sex. The title, *P Q. Test,* was chosen partly to emphasize the fact that these habits are the result of learning, and partly to emphasize the sharp contrast with the I.Q. Experiments with many personality tests, as well as with the *P.Q Test,* have shown that the habits of an effective personality or a socially effective person were in a category quite distinct from the habits of scholastic intelligence and achievement. This fact was confirmed in the nation-wide experiment to validate the *P.Q. Test* (3, 8).

Leadership is a manifestation of social effectiveness. This was assumed when the *P Q. Test* was validated with reference to the criterion, leadership. Leaders, for the purpose of the experiment, were high-school students who had been elected by their classmates to positions of leadership, not leaders appointed by teachers or school officials They were the class officers, officers of the General Organization and other school societies, editors and managers of the school papers, and captains and managers of athletic teams, school band and orchestra, glee club, and other groups. The non-leaders were those who held none of the positions of leadership. The leadergroup was found to have a significantly higher P Q. than the non-leadership group. Therefore, the *P Q Test* measures social effectiveness in terms of leadership.

Leadership

This concept of social effectiveness which we have been developing carries us a step beyond that of social maturity in terms of the status quo. It introduces a qualitative factor, namely: what habits and skills contribute more than others toward social competence and leadership? Accordingly, much of the experimental work with the *P.Q. Test* has concerned itself with item analyses and evaluations of the various habits

and activities. The following guides for the development of leadership and social effectiveness have been deduced from these and other investigations

Social Contact.—Social effectiveness cannot be developed in a social vacuum. One learns the skills of serving others by working and playing with them, not by remaining aloof. Persons who take part in community affairs, or who have been active members of a Y.M.C.A. or Y.W.C.A., dramatic club, the Scouts, or other organized groups, were found to have a higher P.Q. than those who had not. The frequent and often sneering rejection of group activities is eloquent proof that individuals can deliberately choose not to develop the skills of leadership Parents who do *not* believe in letting their children do as they please, but who, by example and discipline, influence them to join the proper groups, are at least beginning their children on the path toward leadership.

Organized Groups.—Highly organized and competitive groups do more to develop leadership than do loosely organized or casual groups. The habits of teamwork and good sportsmanship are the very foundations of social competence and leadership One question of the *P.Q Test* is: “Which of the following have you engaged in frequently in season,—baseball, tennis, bridge, badminton, hockey (a list of 20)?” Those who had engaged in 6 or 7 on this list usually had a higher P Q. than those who had participated in only 1 or 2. The next question is: “In which of these did you practice hard and regularly so that you would become good enough to enter a contest?” Such practice contributes even more toward a high P Q. than does casual participation Among young people the greatest single influence in the development of leadership was intensive participation in competitive games and sports—however, not in one but in several.

Prominent in this list are orchestras, glee clubs, and choirs. Few activities require more intensive practice and team-work than an orchestra Not only must the individual learn to play his own instrument adequately, he must learn to play it in harmony with twenty or thirty others One sour note may destroy the entire effect. The best violin becomes the first violin, a leader, and every competent player may become the leader at times with a solo.

Private Practice.—The practice of some social skills in private is essential to the development of leadership. The person who spends day after day practicing the piano in solitude is developing a skill which may transform self-distrust into social confidence. Almost any skill which one develops to the point of competence may give him or her a sense of confidence and a degree of leadership. Nor does it have to be a skill in music, art, or literature. The more homely arts, such as cooking, knitting, and sewing, telling stories, gardening, being handy with tools, if developed to the point of superiority, can all contribute toward social effectiveness, even to leadership. Nearly every person can acquire some skill to the point where he excels in the ability to interest and serve others.

Economic Independence.—The development of financial independence is indispensable to leadership. The habits of doing socially useful work, as represented by the trait of economic self-reliance in the *P.Q. Test*, are of basic importance. Aside from the experimental and clinical evidences for this conclusion, it seems only natural that the person who leads others and upon whom others depend, should be independent and self-supporting himself.

Opposite Sex.—Social effectiveness includes habits of dealing with the opposite sex. Psychological practice shows that there are some individuals who live in only one half of the social world, either the world of men or the world of women. They have developed emotional and mental barriers which keep them from dealing effectively with members of the opposite sex even in the most simple social situations. Such a barrier is likely to exist whenever an individual has failed to develop, through practice, some of the elementary habits of playing or working with members of the opposite sex. Many of these habits, like social dancing, are obviously of a very specific sort. However, even walking with, talking with, and telephoning to members of the opposite sex are found to be specific habits which must be and can be cultivated by themselves. Specific habits of interesting and serving members of the opposite sex are sharply defined elements in social effectiveness.

Bodily Movement.—Bodily movement is the common denominator of the activities which develop social effectiveness. If one examines the activities mentioned above, or the items which make up the *Social Maturity Scale* and *P.Q. Test*, it will be seen that all of them, whether it be playing the piano alone, playing in an orchestra, working for pay, or playing in a baseball game, involves physical movement. The very word, leader, implies movement and action. The leader is one who moves in advance of his followers, one who must make unusual exertions in their behalf. In terms of our definition of social effectiveness, he has won his position of leadership by his ability to interest and serve others. In so far as he has become the servant of all he may become the master of all.

The common denominator of those pursuits which hinder the development of leadership is the absence of bodily movement, the absence of creative exertion. In this classification fall those time-consuming pursuits in which the individual merely sits and takes things in—the movies, the radio, casual reading, and just plain sitting.

Lead-led.—Leaders must learn to be good followers most of the time. The condition of becoming a leader is the acquisition of skills which make one a good follower, a good teammate. For every situation in which a person acquires these skills to the point of leadership, there are many more situations in which his social effectiveness depends on the skills of being a good follower. A study of leaders and non-leaders (10) in which 40 outstanding leaders and 40 non-leaders were selected in each of three high schools showed that the leaders participated in an average of 6.8 activities compared with 1.7 activities for the non-leaders. The most frequent extra-curricular activities among leaders were competitive athletics. Moreover, the leaders had over twice as much spare time as did the non-leaders, in spite of the fact that they engaged in four times as many extra-curricular activities. The authors attributed this to the fact that non-leaders frittered away much of their time and spent more time in reading and solitary activities whereas the leaders used all their time in a better organized and more efficient fashion. The participation in highly organized activities had brought about a greater integration in their lives generally. "The fact that the leaders participate in games and in attending parties," say the authors at one point, "is highly significant. In games the leader gains more experience with people, and this experience is useful to him in leadership."

Especially interesting is the light cast by this study on the oft-made remark: "We cannot all be leaders, some of us must be followers." This study showed that the leaders were also the good followers, whereas the non-leaders were the poor followers. That is, a leader might be captain or top man in one activity or two, but was merely a member of the team or squad in his other 5 or 6 activities. In other words, the leaders were really followers in 5 or 6 activities, whereas the non-leaders were followers in only 1 or 2, or none. Thus, conspicuous leadership in one activity was based on having become a well-disciplined follower in several activities. The individuals chosen by their fellows to be leaders were tolerated as leaders because they were good followers.

Citizenship

The critical problems of citizenship so sharply illustrated by current perplexities over juvenile delinquency give the above findings unusual importance. In so far as good citizenship can be developed through education, these findings demonstrate the necessity of an almost revolutionary change in the present educational system. This change may be described briefly by saying that many of the compulsory elements of education should be made optional and many of the optional elements made compulsory, that is, the extra-curricular organised group activities should be made part of the regular curriculum for which credit is given, and many of the present required subjects made extra-curricular.

Intelligence and knowledge are necessary but, as we have seen, are by no means sufficient. In fact, there is no correlation today between social intelligence and social effectiveness. Effective citizenship depends on habits of action in organized groups, where most individuals are good followers most of the time but also good leaders some of the time. The individuals who learn the habits of good citizenship in the smaller units such as those enumerated under the heading, organized groups, thereby acquire the basic skills required for effective participation in the larger units of a democracy.

REFERENCES

1. Allport, G. W. and Allport, F. H. *The A–S Reaction Study.* Boston: Houghton-Mifflin, 1928.
2. Doll, E. A. "Annotated Bibliography on the Vineland Social Maturity Scale." *Journal of Consulting Psychology*, IV (1940), 123–132.
3. Link, H. C. *The Rediscovery of Man.* New York: Macmillan, 1937. Pp. 257. Chapters 3 and 4. "The Psychodynamics of Personality and the Habits of Personality," pp. 55–90.
4. Link, H. C. "A Test of Four Personality Traits of Adolescents." *Journal of Applied Psychology*, XX (1936), 527–534.
5. Link, H. C. *Manual for the P.Q. or Personality Quotient Test, 1938 Revision.* New York: The Psychological Corporation.
6. Lockhart, E. G. *Improving Your Personality.* Los Angeles: Walton Publishing Co., 1939. Pp. 512.
7. Moss, F. A., Hunt, T., and Omwake, T. *Social Intelligence Test.* Washington, D. C.: Center for Psychological Service, George Washington University, 1930.
8. Roslow, S. "Nation-Wide and Local Validation of the P.Q. or Personality Quotient Test." *Journal of Applied Psychology*, XXIV (1940), 529–539.
9. Scott, W. D. *Influencing Men in Business, The Psychology of Argument and Suggestion.* New York: Ronald Press, 1911.
10. Smith, M. and Nystrom, W. C. "A Study of Social Participation and of Leisure Time of Leaders and Non-Leaders." *Journal of Applied Psychology*, XXI (1937), 251–259.
11. Strang, R. "Guidance in Personality Development." *Guidance in Educational Institutions*, Chap. VII, 197–228, Thirty-Seventh Yearbook of the National Society for the Study of Education. Part I. Bloomington, Ill.: Public School Publishing Co., 1938.
12. White, W. *The Psychology of Dealing with People.* New York: Macmillan, 1936.

A REPORT ON SCHOLARSHIP EXAMINATIONS GIVEN IN LATIN AMERICAN COUNTRIES FOR THE SELECTION OF STUDENTS TO BE TRAINED IN METEOROLOGY

ALBERT V CARLIN
United States Weather Bureau
and
FREDERIC M. LORD
United States Civil Service Commission

IN AUGUST, 1942, work was begun on the planning of an Institute of Meteorology in Colombia under the joint auspices of the United States Weather Bureau, the Office of Inter-American Affairs, the Defense Supplies Corporation, and the Division of Cultural Relations of the State Department.

The Institute was to award scholarships to 200 Latin American young men for training in Weather Observations, primarily, preparing them for positions in the meteorological services of their own countries or of Latin American Airlines. Courses in meteorology, and refresher courses in physics, mathematics, and English were also to be presented. The Institute was to be staffed partly by U. S. Weather Bureau personnel and partly by Latin American instructors.

The long-range objectives of the Institute of Meteorology in Colombia were threefold.

"1 To increase the safety and regularity of air transportation.

"2. To pave the way for improvements in agriculture through better crop planning and forecasting.

"3. To provide greater facilities for flood forecasting, public health work, and improved operation of public utilities through advance knowledge of weather."

In the planning of the program the question inevitably arose as to how young Latin-American men were to be selected for the training in the Institute. The men selected had to meet the following requirements

"1. They must have a knowledge of physics and mathematics, or they must at least show an aptitude for the study of physical science.

"2 They must have a knowledge of English, or at least have an aptitude for learning languages."

The requirement of a knowledge of English arose from a further aspect of the planning of the Institute. After the six-month course of the Institute was completed, 40 of the better qualified students were to be brought to the United States and provided with advanced meteorological training at one of our universities. Thus, though it was not necessary to select only those who had a good knowledge of English, it was desirable to have as many so qualified as possible, from whom 40 might later be chosen for advanced training in the United States.

The young men were to be selected from each of the Latin American countries, the number from each country being prorated according to population, area, and number of airline-miles.

The first method of selection that presented itself was to base selection on achievement in physics, mathematics, and English, as shown by school records. It was immediately apparent, however, that the scholastic standards of each country are different, and, furthermore, that an evaluation of the courses of study even within each country would be extremely difficult.

It was soon decided that in order to secure objectivity in the selection process, it would be necessary to give the candidates an examination which would measure such factors as knowledge of elementary physics, knowledge of English, and aptitude in physical science and language. The United States Civil Service Commission was asked to design such a battery of tests, to be translated into Spanish, Portuguese, and French.

The examination was scheduled December 15 and 16, 1942, in the capital cities of the various Latin American nations. It was open to individuals between 20 and 30 years of age "who have completed the usual academic preparation for entrance to the national university of their own or a neighboring country."

The written examination consisted of 4 objective tests composed of multiple choice items. Test 1 was a mathematics and physics test composed of eighty items, as follows: 35 mathematical problems in physics, 25 physics information items, and 20 mathematics items. The 20 mathematics items consisted of 9 geometry, 5 algebra, 2 trigonometry, 2 analytic geometry, and 2 calculus items. Test 2 was an English test composed of 100 items, as follows: thirty-five vocabulary items, fifteen grammar items, and fifty reading-comprehension items. Test 3, a language aptitude test, was composed of 150 items, as follows: forty items requiring the competitor to form words from jumbled letters, eighty items requiring the deciphering of coded words, and thirty items requiring the translation of an artificial language. Test 4, a physical-sciences aptitude test, consisted of 100 items, as follows. 25 spatial-ability items, 25 items requiring the competitors to predict the motion of various illustrated mechanical devices, 25 items requiring the insertion of a missing number in a number series, and 25 items requiring the competitors to discover the differences in a set of similar geometrical diagrams.

The examination was printed in Portuguese for administration in Brazil, in French, for administration in Haiti, and in Spanish, for administration in the remaining Latin American nations. The examination scores of 708 competitors were available for statistical analysis. The score on a single test was obtained by dividing the number of correct answers by the number of items in the test and multiplying by 100, the total score was obtained by adding together the number of correct answers on the four tests, dividing by the total number of items, and multiplying by 100.

Findings

Table 1 shows the number of cases, the means, and the standard deviations of total scores for the three translations of the examination. It is to be noted that a comparison of the mean scores of the Haitian competitors, the Brazilian competitors, and the competitors from the Spanish-speaking nations provides no basis for concluding that the three translations of the examination differ sharply in difficulty from each other.

TABLE 1

Means and Standard Deviations of Total Scores on the Three Translations of the Examination

	Spanish	Portuguese	French
Number of cases	553	126	29
Mean	49.5	51.9	53.9
Standard deviation	14.2	12.9	11.6

Table 2 summarizes the obtained data relevant to the effectiveness of the 4 tests as selective devices. The sample of 185 cases upon which the calculations were based was obtained by selecting every third paper from the papers of the competitors who took the Spanish translation of the examination.

The figures of greatest interest in Table 2 are the estimated reliabilities of the four tests. The high estimated reliability of the English Test can be quite plausibly explained on the reasonable assumption that the range in the competitors' knowledge of English was very great, as indicated by the fact that scores on the test ranged from 0 to 96

The extremely high estimated reliability of the Language

TABLE 2

Means, Standard Deviations, Estimated Reliabilities, and Intercorrelations of Scores on the Spanish Translation of the Tests, as Calculated from a Random Sample of 185 Cases

	Test I (Mathematics & Physics)	Test II (English)	Test III (Language Aptitude)	Test IV (Physical-Science Aptitude)
Number of items in test	80	100	150	100
Means of scores	30	50	64	38
Standard deviations of scores	12	19	24	12
Estimated reliability*	.84	.94	.98	.84
Intercorrelations:				
Test I		.36	.53	.48
Test II			.40	.27
Test III				.50

* Computed by the Kuder-Richardson formula, Case IV. See G. F Kuder and M. W Richardson, "The Theory of the Estimation of Test Reliability," *Psychometrika*, II (1937), 151–160.

Aptitude Test cannot be attributed primarily to the length of the test, since reduction of the number of items from 150 to 100 would be expected to reduce the reliability of the test only by about 0.01, nor is it plausible to assume that the range of the language aptitude of the competitors was unusually great, in spite of the fact that scores on the test ranged from 2 to 95, since presumably the range in language aptitude among the competitors is roughly the same as would be found among a group of high-school graduates in the United States One plausible explanation that suggests itself is that many of the competitors may have completely failed to understand the directions of one or more of the 4 different types of items contained in the test If this actually occurred to a large extent, it would be expected that the test would have high reliability as a result of the fact that the test would function as a highly reliable measure of the competitors' ability to understand the particular directions contained in the test. Such a test, however, would be undesirable for the present purpose since not only would the test be a less effective measure of language aptitude than it should be, but also it would not even be a good measure of ability to understand directions other than that particular set of directions contained in the test.

The estimated reliabilities of the Mathematics and Physics Test and of the Physical-Science Aptitude Test are not so high as would be desirable In the case of both these tests the low estimated reliability is presumably attributable to the fact that both tests appear to have been very difficult for the competitors. The ranges of scores on the two tests are from 0 to 70 and from 6 to 74, respectively. For the present purposes, the low reliabilities of these tests are not so serious as might at first appear, since recent developments in test theory indicate that in the case of a very difficult test the scores of the higher-scoring competitors are probably more reliable than would be indicated by the reliability coefficient calculated from the scores of the entire group of competitors

The intercorrelations of the tests are not so high as to indicate a serious degree of duplication among the tests As might have been anticipated, the correlations of the English Test with

the other three tests are the lowest of the obtained correlations, averaging roughly 0.35. The intercorrelations of the other three tests are all close to 0.50, indicating an appreciable, but not too serious, degree of overlapping

* * * * *

Although Latin Americans have limited experience with objective tests, since most of their school examinations are of the essay type, the Latin American officials and the competitors were in general highly pleased that the award of the scholarships was made on the basis of merit determined by objective tests, rather than by personal selection On the whole the project appears to have been highly satisfactory from the point of view of all concerned

---

THE 1943–1944 WORK-SESSION OF THE COUNCIL OF GUIDANCE AND PERSONNEL ASSOCIATIONS

The second Work-Session of the Council of Guidance and Personnel Associations was held at the Hotel Biltmore, New York, on Thursday, November eighteenth The meetings were scheduled for November instead of January with the thought that the representatives of the constituent organizations might profit from an earlier discussion of current problems.

The morning session was largely devoted to the general subject of Rehabilitation.

"Summary of Findings of 1942 Meeting with Emphasis on Changes during 1943"
 Lieutenant C. Gilbert Wrenn, Bureau of Naval Personnel
"Rehabilitation of Service Men and Women"
 Mr. William Gilchrist, Chief of Rehabilitations, Veterans Bureau
 Dr Harry A. Jager, Chief, Occupational Information and Guidance Service, United States Office of Education
"Rehabilitation of Working Students"
 Miss Beatrice McConnell, Director, Industrial Division, United States Department of Labor, Children's Bureau

Gilbert Wrenn gave a brief but very complete summary of the work-session of last January. He pointed out that we are now living out the plans discussed at the last meeting; and we are now anticipating the post-war problems. Mr Gilchrist reviewed government procedure as it is at present in operation. Dr. Jager spoke for the educational program that is just now getting under way. Miss McConnell reported briefly on the dislocation of the school population and the inevitable consequences which will some day have to be dealt with

The afternoon session emphasized the need for the "Preservation of Social Values in a Time of War" The speakers were Mr Robert P. Lane, Executive Director, Welfare Council of New York, and Dr Harry D. Gideonse, President, Brooklyn College Mr. Lane's speech was a philosophic approach to the social values in our civilization. President Gideonse discoursed on the social values which our educational system has tended to emphasize

At the dinner meeting the speaker was Dr Luther Gulick, Chief of the Planning Division, Office of Foreign Relief and Rehabilitation Dr. Gulick had just come from the meetings in Atlantic City and spoke on "Rehabilitation Abroad."

On Friday the Executive Council of the American College Personnel Association held its annual meeting. The members present were E. G Williamson, President; Gilbert Wrenn, Vice-President, D D Feder, Secretary, Thelma Mills and Helen Voorhees Other members in attendance were A. J Brumbaugh, John G Darley, Winifred Hausam, Forrest Kirkpatrick, James McClintock and Robert Moore.

The necessary business was transacted and then the group discussed at length some of the problems of post-war counseling. The results of our deliberation will be reported to you in due time.

Helen Voorhees

MEASUREMENT ABSTRACTS*

Hay, E. M. and Blakemore, A. N. "The Relationship between Clerical Experience and Scores on the Minnesota Vocational Test for Clerical Workers." *Journal of Applied Psychology*, XXVII (1943), 311–315.

The mean score on the *Minnesota Vocational Test for Clerical Workers* classified as experienced was 7.5 points higher than the mean score of inexperienced workers, with a critical ratio of 3.4. Since age, intelligence, and schooling appeared to be unrelated to the difference, it is suggested that either experience tends to screen out the less successful workers or the test is affected by learning. *Lorraine Bouthilet*

Wright, J. H. and Laing, D. M. "The Time Factor in the Administration of the Wonderlic Personnel Test." *Journal of Applied Psychology*, XXVII (1943), 316–319.

In order to test the statement that the time limit on the *Wonderlic Personnel Test* should be ignored because the average individual is penalized by the briefness of the test, a comparison of scores at two different time intervals was made. Although the raw scores for a 24-minute time limit were slightly higher than those for the standard time limit of 12 minutes, the rank order of scores remained almost the same, indicating that for practical personnel purposes the time allowed for the test is adequate. *Lorraine Bouthilet*

Rabin, A. I. "A Short Form of the Bellevue-Wechsler Test." *Journal of Applied Psychology*, XXVII (1943), 320–324.

Correlations of the combined scores on three sub-tests (Comprehension, Arithmetic, and Similarities) of the *Bellevue-Wechsler Test* with the total test and the *Army Alpha* yielded correlations as high as those usually obtained between two intelligence tests. The subjects in the study were 92 nurses and 200 hospital patients. In using the short form the I.Q. is determined by administering the three tests in the usual manner, computing the score by the formula, $X = \frac{C\,A\,S}{3} \times 10$, in which X is the total score and *C A S* the weighted scores in the sub tests, and referring to the I.Q. tables for the various age groups. *Lorraine Bouthilet*

Lawshe, C. H., Jr. and Thornton, G. R. "A Test Battery for Identifying Potentially Successful Naval Electrical Trainees." *Journal of Applied Psychology*, XXVII (1943), 399–406.

The purpose of this study was to develop a battery of tests to be used in identifying men most likely to be successful in a Naval Training School for Electrician. From correlations of scores on a preliminary group of tests with grade-point averages in the five courses, a final battery of three tests, ability to read simple measurements and solve arithmetic problems, electrical information, and mental alertness was selected. Using a regression equation to predict achievement of a new group of men, it was possible to establish a critical score, on the basis of the grade-point averages predicted from the tests, which would eliminate practically all failures. *Lorraine Bouthilet.*

---

* *Edited by Forrest A. Kingsbury*

Gottsdanker, R. M. "Measures of Potentiality for Machine Calculation." *Journal of Applied Psychology*, XXV (1943), 233–249.

A battery of nine paper and pencil tests, chosen on the basis of a job analysis, was assembled in order to obtain a measure of the abilities needed for proficiency in the operation of machine calculators. Scores of the battery were validated against the performances of 44 women learning to be machine operators. As a result of an analysis of the validity correlations, a final battery of three tests, the *Number-Dot Location*, *Arithmetic Computation*, and *Tapping* tests, selected by means of the Wherry-Doolittle test selection method, gave a multiple correlation of .57 with the criterion. The test retest reliability coefficient for test scores from another group of subjects was estimated to be .89 for the final battery. The author believes that the multiple correlation coefficient of validity may be spuriously low due to the homogeneity of the criterion group and the errors of measurement in their achievement scores. *Lorraine Bouthilet.*

Frandsen, A. N. and Hadley, J. M. "Prediction of Achievement in a Radio Training School." *Journal of Applied Psychology*, XXV (1943), 303–310.

In an attempt to predict success in a radio training school for Naval and Marine recruits, the correlations between tests of intelligence, interest in radio, mathematical ability, technical information of various kinds, and the achievement scores on the curriculum were estimated. It was found that the most efficient measure for predicting achievement by use of the multiple regression equation was a combination of either one of the mathematics tests and the *Technical Information in Electricity Test*, with scores on the intelligence test and interest questionnaire adding little to the accuracy of prediction. *Lorraine Bouthilet.*

Maddy, Nancy Ruth. "Comparison of Children's Personality Traits, Attitudes, and Intelligence with Parental Occupation." *Genetic Psychology Monographs*, XXVII (1943), 3–65.

The most significant results of this study of the emotional make-up and intelligence of children in two widely divergent socio-occupational groups indicate that reliable differences in intelligence and personality traits do exist between children in these two groups, with greater differences being found for girls than for boys. Certain traits, extroversion, dominance, and emotional stability, were more pronounced in children of the professional group whereas children of the semi-skilled seemed to have more worries. Variations from the average scores for the particular group to which the children belonged were noted in the cases of those children who lived in neighborhoods not representative of their occupational groups. *C. A. McNally*

Ames, Viola. "Factors Related to High School Achievement." *The Journal of Educational Psychology*, XXXIV (1943), 229–237.

The author is concerned with the relation of certain personality traits which may be used in addition to intelligence scores in the prediction of scholastic achievement in high school. The results of the study indicate a significant relation between the ability to conform to school situations and scholastic achievement, whereas no relation was found between achievement and the ability to succeed socially. *C. A. McNally*

Miller, Leo R. "Some Results of Retesting Elementary-School Pupils with the Stanford Revision of the Binet-Simon Scale." *Journal of Educational Psychology*, XXXIV (1943), 237–242.

Forty five boys and forty five girls of one grammar school were tested on the 1916 *Stanford-Binet* and retested on the 1937 *Revised Stanford-Binet* (Form L). The mean chronological age at the time of the first test was 5.86 years and at the time of the second test 12.3 years, the mean interval between the two tests being 6.5 years. Variations in scores between the two tests ranged from 0–34 I.Q. points. The author found that boys are more likely to gain in I.Q. score over a long period of time, whereas girls are more likely to lose. He concludes from his results that prediction based on one test, given at the beginning of the child's school career, without *retesting*, can be more harmful than failure to give any intelligence test at all. *Miriam Rotman.*

Gilkinson, Howard. "The Seashore Measures of Musical Talent and Speech Skill." *Journal of Applied Psychology*, XXVII (1943), 443–448.

The proposition that vocal performance is dependent upon some types of auditory abilities that are measured by the *Seashore Tests* is investigated. Conclusions, based on the results of three studies, are: (a) No positive relationship can be determined between specific speech qualities and the *Seashore Tests*. (b) Slight positive relationship was found between degree of speech excellence and the score on the *Seashore Tests*. (c) "Practical diagnostic value in general speech courses is very questionable." *Francis Madland.*

Teegarden, Lorene. "Occupational Differences in Manipulative Performance of Applicants at a Public Employment Office." *Journal of Applied Psychology*, XXVII (1943), 416–438.

The author has previously described a battery of performance tests which were administered to applicants at the Cincinnati Employment Center and from which norms for each sex were developed. These tests were grouped together to get some measure of manipulative ability. In this study each applicant's performance on the battery is studied in relation to that of other applicants who had had experience in the same occupation or occupations as those claimed by the applicant. Each occupational group, taken as a whole, includes records of long and short experience, of efficient and inefficient workers. All ages were included, with the majority between sixteen and twenty-five years of age. Only white subjects were used. The occupations of the men were (1) helpers in skilled trades, (2) truck drivers and chauffeurs, (3) truck helpers and loaders, (4) factory labor, hand operators, (5) factory machine operators, (6) sales clerks, (7) waiters, kitchen helpers, bus boys, and dish washers, (8) manual laborers; those of the women were (1) factory workers at hand operations, (2) assemblers, inspectors, testers, (3) packers and wrappers, (4) sales clerks, (5) waitresses, (6) domestic workers. The results show that differences do appear between certain occupational groups. These differences are greater in ability to solve problems, in accuracy of movement, and in ability to react to multiplicity of details than in rapidity of hand movement or coordination of two-hand movement. *Miriam Rotman.*

Goodman, Charles H. "A Factorial Analysis of Thurstone's Sixteen Primary Mental Abilities Tests." *Psychometrika*, VIII (1943), 141–151.

This article deals with a factorial analysis of Thurstone's *Primary Mental Abilities Tests*. The analysis was made in order to determine whether the tests designated to measure a particular ability would be found upon completion of the analysis to be clustered together with significant loadings and be isolated from the tests of the other abilities. The results show that most of the tests specifically designed to measure six of the abilities were isolated with loadings varying in size. The tests for the remaining ability were not isolated. It is also shown that the tests are complex in nature, measuring more than any one single ability. (Courtesy of *Psychometrika*.)

Festinger, Leon. "An Exact Test of Significance for Means of Samples Drawn from Populations with an Exponential Frequency Distribution." *Psychometrika*, VIII (1943), 153–160.

The mathematical derivation of a test for determining the fiducial limits of, and significance of difference between, means when the samples are drawn from exponential populations is presented. The test for differences between means takes the particularly simple form of the *F* test (the ratio of the larger to the smaller mean) with each mean possessed of $2n$ degrees of freedom, n being the number of cases in the sample. Random sampling, a range of scores upwards from a lower limit of zero, and independence of means from each other are necessary assumptions for the use of the test. Examples of situations in which the test should be used are given, together with a description of the necessary computational procedures. Comparisons of the results of the application of this test with the erroneous application of the critical ratio on actual data show that rather large discrepancies exist between the two tests. Results obtained by applying tests which assume normality to exponential distributions are subject to much error. (Courtesy *Psychometrika*.)

Mosier, Charles I. "On the Reliability of a Weighted Composite." *Psychometrika*, VIII (1943), 161–168.

A general formula for the reliability of a weighted composite has been derived by which that reliability can be estimated from a knowledge of the weights whatever their source, reliabilities, dispersions, and intercorrelations of the components. The Spearman-Brown formula has been shown to be a special case of the general statement. The effect of the internal consistency or intercorrelation of the components has been investigated and the conditions defining the set of weights yielding maximum reliability shown to be that the weight of a component is proportional to the sum of its intercorrelations with the remaining components and inversely proportional to its error variance. (Courtesy *Psychometrika*.)

Guttman, Louis and Cohen, Josef. "Multiple Rectilinear Prediction and the Resolution into Components II." *Psychometrika*, VIII (1943), 169–183.

Given a battery of n tests that has already been resolved into r orthogonal common factors and n unique factors, procedures are outlined for computing the following types of linear multiple regressions directly from the factor loadings: (i) the regression of any one test on the $n-1$ remaining tests, (ii) all the n different regressions of order $n-1$ for the n tests, computed simultaneously; (iii) the regression of any common factor on the n tests, (iv) the regressions of all the common factors on the n tests computed simultaneously; (v) the regression of any unique factor on the n tests, (vi) the regressions of all the unique factors on the n tests, computed simultaneously. Multiple and partial correlations are then determined by ordinary formulas from the regression coefficients. A worksheet with explicit instructions is provided, with a completely worked-out example. Computing these regressions directly from the factor loadings is a labor-saving device, the efficiency of which increases as the number of tests increases. The amount of labor depends essentially on the number of common factors. This is in contrast to computations based on the original test intercorrelations, where the amount of labor increases more than proportionately as the number of tests increases. The procedures evaluate formulas developed in a previous paper. They are based essentially on a shortened way of computing the inverse of the test intercorrelation matrix by use of the factor loadings. (Courtesy *Psychometrika*.)

Greenwood, J. A. "A Preferential Matching Problem." *Psychometrika*, VIII (1943), 185–191.

A matching method proposed by Dr. C. E. Stuart is presented in some detail and the essentials for a test of significance are derived. The method differs from the older matching methods in that partial credit is allowed for a near miss. A slight variation of the method permits the matching of one item with M sets of n traits. (Courtesy *Psychometrika*.)

NEW TESTS*

Cooperative Pre-Flight Aeronautics Tests Provisional Form A, 1943
Test 1 *Aerodynamics and Aircraft Structures*, by Col R. H Drake, Ens. K. V Jackman, Alexander Joseph, Louis Teichman, and Carl A Pearson. Time, 40 minutes.
Test 2: *Aircraft Engines*, by Col. R. H. Drake, Ens K. V Jackman, James L. G Fitzpatrick, and Carl A. Pearson Time, 40 minutes.
Test 3: *Meteorology*, by Col R H Drake, Ens. K. V Jackman, and Rosalie K. Kuschen. Time, 40 minutes.
Test 4: *Navigation*, by Col. R. H. Drake, Ens. K. V Jackman, and Meta Rothholz. Time, 40 minutes.
Test 5: *Part I, Radio and Communications, Part II, Civil Air Regulations*, by Col. R. H Drake, Ens. K. V Jackman, and Alexander Joseph. Time, 40 minutes.

These tests are for the 11th and 12th grades and junior college level. Norms for total and part scores for the 12th grade are furnished. Each booklet, 10 to 99 copies, 6¢; 100 or more copies, 5½¢ Specimen set, including one copy of each booklet, 50¢. The tests may be hand- or machine-scored machine-scored answer sheets, 1½¢ each. Machine-scoring stencil, 30¢. Published by the Cooperative Test Service, 15 Amsterdam Avenue, New York, N Y

Cooperative College Biology Test, Provisional Form T, by Carl A. Pearson and others, 1943 Time, 90 minutes. This test is for first-year college biology classes, for which tentative percentile norms are provided. 10 to 99 copies, 7¢ each; 100 or more copies, 6½¢. Specimen set, 25¢ The test may be hand- or machine-scored machine-scored answer sheets, 2¢ each, machine-scoring stencil, 30¢. Published by the Cooperative Test Service, 15 Amsterdam Avenue, New York, N Y

Cooperative College Mathematics Test, Provisional Form T, by Emma Spaney and T Freeman Cope, Marguerite D Darkow, Richard E. Henry, C. V Newsom, Fred Robertson, and E R Smith, 1943 Time, 40 minutes This test is for first-year college mathematics classes. 10 to 99 copies, 6¢, 100 or more copies, 5½¢ Specimen set, 25¢ The test may be hand- or machine-scored machine-scored answer sheets, 1½¢ each, machine-scoring stencil, 15¢. Published by the Cooperative Test Service, 15 Amsterdam Avenue, New York, N Y

Watson-Glaser Tests of Critical Thinking, Form A· Battery I, Discrimination in Reasoning, Battery II, Logical Reasoning, by Goodwin Watson and Edward Maynard Glaser, 1942 No exact time limit, but the time required is about one hour, which may be divided into two sittings This test is for high school, college and adult levels. Package of 25 tests, either battery, $2.00. Specimen set, both batteries, 45¢ This test is hand- or machine-scored machine-scored answer sheets, package of 25 for either battery, 70¢ Published by World Book Company, Yonkers-on-Hudson, New York, N Y.

Iowa Silent Reading Tests, Elementary Test, Forms Am and Bm revised, Forms Cm and Dm new edition, by H A. Greene and V H Kelley, 1943. Time, 49 minutes. This test is for grades 4 to 9, for which norms are provided. Per package of 25, $1.45 Specimen set, 35¢ The test may be hand- or machine-scored machine-scored answer sheets, per package of 25, $1 15, scoring keys for any form, each 60¢ Published by World Book Company, Yonkers-on-Hudson, New York, N Y

---

* Prepared by Mrs. John Segrave, Jr., The University of Chicago.

Iowa Silent Reading Tests, Advanced Test, Forms Am and Bm revised, Forms Cm and Dm, new edition, by H A Greene, A N Jorgensen, and V H Kelley, 1943. Time, 45 minutes This test is for high-school and college students, for which norms are provided. Per package of 25, $1.80 Specimen set, 40¢ The test may be hand-or machine-scored machine-scored answer sheets, per package of 25, $1.15, scoring keys for any form, each 80¢ Published by World Book Company, Yonkers-on-Hudson, New York, N Y

Victory Corps Aeronautics Aptitude Test U S. Office of Education, Federal Security Agency Grades 10, 11, 12 Time, 90 minutes. Test booklets, $7.50 per 100; answer sheets, $2.00 per thousand. For sale by Superintendent of Documents, Washington, D C. Sample copies may be obtained from the U S. Office of Education, Federal Security Agency, Washington, D. C

Personnel Test, by E F Wonderlic. Forms A and B For adults. Time, 12 minutes. $5.00 per 100, specimen set (10 copies), 75 cents. Published by E. F Wonderlic, 919 North Michigan Avenue, Chicago, Illinois.

Inventory of Affective Tolerance, by Robert I Watson and V E. Fisher $4.00 per 100, $2.25 per 50, specimen set, without scoring keys, 15 cents; scoring keys, 25 cents per set. Published by Sheridan Supply Company, P.O. Box 837, Beverly Hills, California

NEWS NOTES*

In July a new Katharine Gibbs school was opened in Chicago. Dorothy Coats, who had formerly been Placement Director in the Boston school, was made Director of the Chicago school

Nora A. Congdon, Colorado State College of Education, has been in Washington on a temporary assignment with the Civil Aeronautics Administration as Educational Consultant, Aviation Education Service. The C.A.A. in co-operation with the American Council on Education is making a study evaluating certain phases of aeronautics courses as taught in high schools during the year 1942–43

Lt. Colonel Lyle W Croft is Commanding Officer of SCU No. 3410 Star, Basic Training Center, ASTP, Fort Benning, Georgia. This unit is the classification, testing, and statistical center for processing Army Ground Forces Personnel Colonel Croft says the work reminds him very much of the work of the admissions office and the personnel office of a large university They have complete interviewing, testing, reviewing, screening, and statistical teams These teams are formed for specialized work so that they can completely classify and screen large numbers of candidates per week. They also have a Personnel Consultant Section which handles problem and maladjustment cases. It is Colonel Croft's belief that the personnel work being done in the Army will affect college personnel work in new trends and developments.

On October first, Neal E. Drought resigned from his position as assistant dean and assistant professor of education at Hamilton College, to join the staff of the Personnel Planning and Research Division of R.C.A.-Victor in Camden, New Jersey His new responsibilities involve the development and coordination of personnel policies and procedures in the several R.C.A.-Victor plants

Wm. C. Johnson, Jr., Director of Placement, Virginia Polytechnic Institute, has accepted a position with the Goodyear Aircraft Corporation in Akron, Ohio.

Harry W Seamans, General Secretary of the Penn State Christian Association, Pennsylvania State College, has recently accepted a position as Senior Community Participation Adviser, National Housing Agency, Office of the Administrator, Washington, D. C.

Since early in 1942, Donald E. Super, Associate Professor of Educational Psychology and Director of the Personnel Bureau at Clark University, has been on leave of absence for duty with the armed services. At present he is a Captain in the Army Air Corps stationed at Nashville, Tennessee, with the Nashville Army Air Center (AAFCC) Psychological Research Unit No. 1 As Test Development Officer for this unit his duties include job analysis, and development of tests for pilots, navigators, and bombardiers, and have involved such things as field studies, flying on training missions, and going partially through pilot training as a participant observer. He has a number of enlisted psychologists as assistants in the Test Development Section.

---

* News items concerning members of the American College Personnel Association should be sent to Grace E. Manson, Northwestern University, Evanston, Illinois.

Lieut. (j.g.) R L. Swann, U.S.N R., was granted a leave of absence as Director of Personnel at Green Mountain Junior College, to enter the Navy as a reserve officer After serving seven months as Psychologist at the U S Naval Aviation Selection Board, Dallas, Texas, he was ordered to the U S Naval Flight Preparatory School, University of South Carolina, Columbia, South Carolina, where he is Psychological Officer Much of his time is spent in counseling the V-5 cadets on study habits and personal problems

J E Walters, Vice President in charge of Personnel and Labor Relations, Revere Copper and Brass, Inc., has resigned to join McKinsey and Co., management consultants, 60 E. Forty-second St., New York City

Public announcement has recently been made by the War Manpower Commission of revised policies and procedures in connection with the placement and transference of personnel, including professional and scientific personnel registered with the National Roster These new regulations are embodied in Regulations 4 and 7 of the War Manpower Commission and Selective Service Local Board Memoranda 115, 115-B, and 149

AN INVENTORY OF STUDENTS' GENERAL GOALS IN LIFE

HAROLD B DUNKEL[1]

Cooperative Study in General Education

WHAT basic beliefs now constitute college students' "philosophies of life" or "designs for living"? What does the college student consider the main goals of his life? For the sake of what values does he think that he lives? To what extent does his total pattern of values seem to meet certain criteria of a "good" design for living? The attempt to answer these and similar questions led to the project in philosophy of life and religion undertaken by the Cooperative Study in General Education[2] Almost without exception the colleges stated some such institutional objective as "to aid the student in developing a useful and desirable philosophy of life," and faculty members insisted that their institutions did not merely state this objective in the catalogues but were sincerely striving by various means to aid the student in developing an adequate personal philosophy

Four possible steps in dealing with the problem of students' philosophies of life can be indicated by the following questions I *Do and should students have philosophies of life?* II. If so, what are these philosophies? III Once we have ascertained a student's philosophy, do we (the student and the faculty) consider it satisfactory, *and on the basis of what criteria do we* decide? IV If a student's philosophy appears unsatisfactory, how can the college assist the student in working out a satisfactory view of life?

Our colleges believed that the answer to the first question was "Yes," that students should have a philosophy. Many teachers and institutions had made a practice of having students write essays setting forth their philosophies Since these essays had all the limitations of any verbal statement (they are what the student *thinks*, or *wishes to think*, or *wishes the reader to think*, is his philosophy), these teachers were of course not under the misapprehension that there is a perfect correspondence between the way men live, and the way men say they would like to live The hypocrite, the person with selfish or anti-social aims which he fears to express, the victim of social or economic pressure, all undeniably exist Those undertaking the project believed, however, that it is equally incorrect to assert that there is *absolutely no relation* between the way in which people talk about life and the way in which they live. The group felt that in many cases the verbal statement about the kind of life the student was seeking would be a useful index, even though not a perfect one, of the kind of life the student was living or seeking to live.

In the case of many students, when there is a discrepancy between stated beliefs and conduct, this inconsistency exists *without the person's being aware of it. For these students*, an opportunity to state their beliefs precisely and to examine the implications of those beliefs for action may often lead the student to harmonize more closely his pattern of beliefs and his pattern of living.

Other students actually do not know what they believe. They have never thought much about their way of life or about what they believe. Having acquired their schemes of value rather unconsciously and at random from various sources, these students, when they leave the pattern of life into which they were born and grew up, and when on coming to college they meet new and perhaps conflicting values represented by new environment, new friends, and new experiences, have considerable difficulty in ordering their personal lives for the first time.

Without help in seeing and resolving some of this conflict and confusion, the student may go through much if not all of his adult life, a victim rather than a master of these warring values The purpose of the proposed device was, therefore, to enable the student to see clearly what he says he believes, to compare it with the philosophies of other students, and to obtain help in overcoming difficulties and conflicts in his own position.

The group did not believe that all these results could be satisfactorily produced at the verbal level alone. Even on the verbal level, we certainly did not wish that a student should work out for himself, once and for all, an immutable philosophy of life We believed, however that

> most teachers . will agree that no student should merely drift through life, allowing his major decisions and actions to be determined for him entirely by circumstances—to be merely the resultant of external forces impinging upon him Each student should bring to every important life decision some "sense of values" which is his own . . Of course such a "design for living" can never be completely explicit . . In a brief, oversimplified, presentation it [the idea of a design] sounds unduly rationalistic—as if each student were expected to chart all the details of his life, to be conscious of all the reasons for his every act, and to refer all decisions to some grandiose notion of life and the universe We do not mean this[3]

Rather the philosophy should be a "working hypothesis" for living, to be revised as and if necessary, in the light of new situations and insights Though a satisfactory verbal statement of a philosophy of life is not the whole journey, it is at least a step in the right direction. Such reasoning as this, then, led our group to the belief that students should have a philosophy of life (the group's answer to the question of Step I) and to an interest in securing a verbal statement of this philosophy (Step II).

A further result of conferences between faculties and staff was the decision that some *objective* means of securing and recording the student's statement of his philosophy was desirable. Although, to be sure, the response in the form of an essay gives certain insights which probably cannot be made available through objective devices, the existence of the objective instrument would not preclude the use of the essay

On the other hand, the objective instrument would have, *the group felt, several advantages In the first place, the* "philosophies" thus recorded would be more exactly and more easily comparable. For example, the score of one student could then be compared to his own scores at other points in his educational career, to the scores of other individuals, and to the scores of groups. Similarly, scores of sections, classes, or entire student-bodies could be compared, and if desired, judged. Likewise the philosophy of a student who, when left to himself, deals in airy generalities would have more points of comparison with that of a student who plunges into petty minutiae; and the statement of the verbally facile student could more meaningfully be compared with that of one less articulate.

Furthermore, the records of these philosophies would be more concise and more usable. Faculty-members who wished to learn about the philosophy of a group of students would no longer face the difficult and often impossible task of reading through hundreds of pages of student manuscript. Rather, the results stated numerically would permit fairly rapid summarization and comparison. Then finally, instead of existing like the essay in a single copy tucked away in some personnel-folder or teacher's files, copies of the results of an objective record could be made available for study and consultation, easily and inexpensively, to every person concerned with the student For these reasons, then, the preparation of some objective device of this sort seemed eminently desirable as a first step

Many aspects of the student's philosophy of life deserved study; but, for various reasons which need not be discussed here the group began work by studying the main goals which students hold. Hence, though other studies were planned to supplement the information gained from this device, the first instrument which the group undertook to construct was the "Inventory of General Goals of Life."

---

[1] Work on this project has been carried on by faculty members of colleges participating in the Cooperative Study in General Education and by members of the central staff working in the field of Humanities, George E. Barton, Jr, Walker H. Hill, and the author Dr Barton, now Lt. Barton of the U S. Air Corps, directed the original organization of the project and the preparation of the first form of the inventory The complete report on the project, including norms and data on reliability and validity, is being published this summer

[2] The Cooperative Study is an organization of approximately a score of colleges who attack cooperatively, with the aid of a central staff, their common educational problems. For a report of the Study's work, see Ralph W Ogan, "The Cooperative Study in General Education," *The Educational Record*, XXIII (October, 1942), 692–703 Those interested in greater detail about the development of this project, should consult the *Staff News Letter* (The Cooperative Study in General Education, 5835 Kimbark Ave., Chicago, Illinois), vol II, no 6

[3] *Staff News Letter*, Vol. II, no. 6, p 6.

First, a list of possible main goals was secured from papers written by students and from teachers in the group. In the selection of the goals finally included in the inventory, two criteria were dominant The first was that the goals listed, though couched in student phraseology, should give an adequate representation of certain of the great historical traditions of philosophy and religion The second was that the list should also include goals which, though less common in formal philosophy and religion, are familiar in our culture or in "cracker-barrel philosophy." Within the limitations imposed by practical conditions, we wished each student to find expressions which he could consider adequate statements of elements of his own point of view.

The number of possible "philosophic" positions (to use the term in its widest possible sense) and the number of statements that can be framed to express any single one are, of course, enormous. Yet the practical situation demanded that the final list be extremely brief. We tried to secure brevity in three ways. (1) We allowed certain goals between which distinctions are customarily made in philosophy and which yet appear to be closely related to one another in the minds of most students, to stand together in a single statement[4] (2) Certain goals such as the attainment of Nirvana, which are familiar in the history of thought but which are not commonly held by American college students, were omitted (3) Certain goals common in student philosophies but in a subordinate position were also omitted[5]

Closely connected with the selection of the goals was the question of what technique should be used. Since the purpose was to get the student to rank the goals in order of their importance and acceptability to him, the device of "paired comparison" appeared appropriate since this technique facilitates accurate and easy ranking of various possibilities. Thus each

---

[4] An example of this sort of statement of goal is "Peace of mind, contentment, stillness of spirit."

[5] An example is "good health" Though many students list it as a goal, few if any students seek health as an end in itself Rather they consider it a necessary condition for attaining other goals which they consider more important. Since the aim of the inventory was to secure some indication of the more dominant goals, these subordinate goals were omitted.

of the twenty goals is paired once with all the other nineteen, and the student is asked to choose one goal of each pair As a result of this process, the goal for which the student has the greatest preference will be selected nineteen times (i e., every time it appears) or will have a "score" of nineteen The student's second choice will be selected eighteen times (a score of eighteen), being rejected only in the pair where it appears with the most acceptable goal. The other goals follow with diminishing scores until the goal which the student finds least acceptable or rejects most emphatically is reached In short, the student who manages to make his choices with perfect consistency will give to goals "scores" which will rank them in order.[6]

In interpreting the inventory, one looks first at the goals which are ranked at the head of the list (The exact number of goals considered depends partly on the nature of the particular goals concerned, partly on the arrangement of the scores in terms of gaps or ties in the ranking.) These goals at the head of the list are the statements which most appeal to the student Next the goals at the foot of the list are considered since the student's philosophy as indicated by the goals most readily accepted is often further defined and clarified by the goals which he rejects (or accepts least readily)[7]

The way in which the inventory records a philosophy of life can probably best be indicated by an example This student *before* taking the inventory, wrote a brief essay stating her philosophy This essay is reproduced here and followed by a

---

[6] For several reasons which are too complicated for discussion here, perfect "consistency" is rare, and ties are common in the rankings made by students

[7] In interpreting the inventory, usually little weight is given to those goals which appear in the middle of the list. The reasons for this procedure are both philosophical and statistical.

Statistically, extreme deviations are of course much less likely to occur because of pure chance

Philosophically, the basic concepts of many positions can be stated by accepting relatively few of the statements in this list and by rejecting a few others. For a particular philosophical position, many of these goals listed may be irrelevant or meaningless. The form of the test does not, however, permit the student to discard these goals literally He must continue to make choices involving them He tends to mark those goals to which he is indifferent somewhat lower than those goals which he accepts as statements of his position, yet somewhat higher than those goals which his position necessarily rejects. In other words, the goals toward which he is neutral or indifferent tend to appear in the middle of his ranking.

listing of the order in which this student subsequently ranked the goals of the inventory. The case is fairly typical in that, while the statements in the philosophy are not verbally identical with the wording of the goals in the inventory, the general nature and tone of the philosophy are reproduced

My Philosophy of Life

I find it difficult to state my philosophy of life One's philosophy is far from being an immutable thing, and mine is constantly changing from contact with the outside world. The best I can do is to state the basic principles which are elementally stable.

I first of all think that life is to be enjoyed Happiness no matter how brief, leaves a lasting impression that lives on from one period to the next Its memory fills the more somber moments with a hope for better things to come If I am happy I can be successful, and success gives the impetus to do better things.

Every man on earth is here for some purpose. I believe that I can attain that purpose by doing the best I can with everything I do. The thing at which I was intended to excel will show itself in the enjoyment I derive from doing it At the present time my greatest source of joy is in writing I cannot say yet that writing is my talent, What I write is immature and sometimes I become very discouraged when my words do not match my thoughts But to know that I have written a line full of life and color gives me more pleasure than any other thing I do

I am not concerned with the life to come after this one. I do not pretend to know what it holds for me or how I can change it. I *am* concerned with the present It is of too much importance to be considered as a doormat to the house of the next life. The present is a house in itself and it must be lived in. I believe that if this life were not meant to be enjoyed it would not hold so many sources of joy

It is not my ambition to be remembered for generations to come because of some great accomplishment If I succeed in doing one thing really well, I shall be satisfied. It was never intended that I be a famous person, but I am going to see to it that I am a successful one.

Inventory Scores

"Score"	*Goal*
18	Getting as many deep and lasting pleasures out of life as I can
17	Promoting the most deep and lasting pleasures for the greatest number of people
17	Self-development—becoming a real, genuine person.
16	Fine relations with other persons
15	Making a place for myself in the world, getting ahead
14	Handling the specific problems of life as they arise.
13	Peace of mind, contentment, stillness of spirit.
11	Power; control over people and things.
10	Serving the community of which I am a part
9	Self-sacrifice for the sake of a better world.
9	Living for the pleasure of the moment
7	Serving God, doing God's will
7	Achieving personal immortality in heaven.
6	Self-discipline—overcoming my irrational emotions and sensuous desires
6	Doing my duty.
6	Survival, continued existence.
4	Being able to "take it"; brave and uncomplaining acceptance of what circumstances bring
3	Finding my place in life and accepting it.
2	Realizing that I cannot change the bad features of the world and doing the best I can for myself and those dear to me.
1	Security—protecting my way of life against adverse changes

In her essay, this girl first mentions, as an aim in life, the desire to be happy This desire is reflected in the goal which heads her list in the inventory, "Getting as many deep and lasting pleasures out of life as I can" (Of the statements in the inventory this is the expression nearest to "happiness") The rank she gives to "Peace of mind" indicates the same attitude The social aspect of this same aim appears in two other goals she ranks high. "Promoting the most deep and lasting pleasures for the greatest number of people" and "Fine relations with other persons"

The second aim emphasized in the essay is the desire to be successful, to do the best she can in everything In the inventory this aim is indicated by the high ranking of the goals of "Self-development" and "Making a place for one's self in the world"

In her essay, this student stresses the present and is specifically indifferent only to the problem of "the life to come after this one." Hence the goal of "Handling the specific problems of life as they arise" ranks high in her list and that of "personal immortality" falls in the lower middle

At the end of her list come, as might be expected in the case of one who wishes to make a place for herself in the world, the more passive goals of "Finding my place in life and cheerfully accepting it" and merely "Being able to take it"[8]

---

[8] Taking Steps III and IV, one should ask whether this philosophy as expressed is adequate, and if not, how the institution can assist the student in securing a better point of view. Although considerable work has been done along this line, limitations of space prevent the inclusion of any account of it in the present article.

Possibly these brief comments will suffice to show the more important relations between the essay and the inventory, though practice gained through interpreting a number of scores makes the results more meaningful to the user Some differences between essay and inventory may also be noted For example, the interest in writing, stressed in the essay, does not of course appear in this inventory On the other hand, the inventory reveals the student's point of view on a number of issues which she did not mention in her essay Unfortunately a single example cannot illustrate the many varieties of viewpoint and the striking contrasts between them which can be indicated by the inventory, nor can it show the analyses and comparisons possible for the scores of groups of students[9]

It is still too early to make a final evaluation of the inventory Data are still coming in, and studies of the validity and other aspects of the instrument are still being conducted by the staff and by the cooperating institutions. Nonetheless it is fair to say on the basis of the evidence now available that the inventory has proved extremely useful and seems a reliable instrument for the purpose for which it was intended

---

[9] Studies of several groups of students and adults can be found in *Staff News Letter*, vol IV, no. 11

MAJOR STRATEGY VERSUS MINOR TACTICS IN MERIT ADMINISTRATION

FRED S BEERS AND CECIL E. BROLYER
Social Security Board

Civil service or merit systems are relatively new to the American scene of government Even the most venerable of them are only two generations old, but between 1935 and 1944 the number skyrocketed until there were state-wide systems or ones that included a combination of departments in all but three states Stimulation for this phenomenal growth came in part from federal legislation, particularly the Social Security Act, which in 1939 was amended so that selection of personnel on a merit basis was one of the requirements for receiving grant in public assistance and unemployment insurance

Passing laws is risky business—witness the prohibition amendment. Civil-service legislation has sometimes been passed in nearly as precipitous a fashion Although the results have been notably better, they have left little doubt that statutes alone, indispensable as they are, cannot make a merit system Practice must keep step with legal precept if public acceptance is to be sufficiently durable to make laws effective. Philosophically inclined members of the legal profession say that 80% of the people must be doing voluntarily what a law dictates if the 20% who need coercion are to be brought into line.

For more than a century the American tradition of personnel selection was opportunistic. Officeholders were dispossessed regularly with each change in administration. Having an average expectation of tenure amounting to two years, they spent the first year learning what they were supposed to do and the second helping to re-elect the chief administrator This practice was a main source of popular merriment The

cost was enormous, but Americans have always paid handsomely for their entertainment. In the field of business and industry, competitive advantage, often cut-throat in nature, took precedence over other considerations Extended to persons, the principle found expression in hiring, firing, and fierce competition for jobs

Even in education the pressures against reasonable selection were legion, and there had been marked lag in the adjustment of opportunity to individual and social needs For example, before the College Entrance Examination Board was founded in 1900 the colleges had been "evaluating" secondary schooling by means of records not even superficially comparable. When they had muddled through to elements of a rational policy for admission, some presidents still objected One, more naive or pragmatic than the others, "pointed out that a college might wish to show special favor to sons of large benefactors, to sons of trustees, or to sons of public men of importance who presumably would have difficulty in meeting the announced admissions requirements"[1]

Eventually a system of personnel selection may develop that can with justice be called scientific Those who have studied the problem know that on theoretical grounds an infinite number of such systems exist as possibilities. Two questions would have to be answered if a particular system were chosen for practical application· Which of the possible systems would be *most useful*, and *Who* would select it? According to democratic ideology, the citizens would decide and choose. It would be useful to the extent that its results were sensed as beneficial to the entire society, it would be selected by the entire society. If the urge for scientific selection were compelling, we would now be searching for our undefined or basic terms; we would be exploring the postulates smallest in number but *useful* to attaining our aims. With these goals reached, our conclusions and their application would follow as inevitably as night follows day But generations may come and go before public opinion can forge a method of selection worthy to be called scientific.

---

[1] *The Work of the College Entrance Examination Board, 1901–1925.* New York Ginn & Co., 1926.

Meanwhile we must consider available alternatives Among these the merit system seems best, although its imperfect techniques can be much improved as this article will attempt to point out

In their present stage of development merit systems cannot pretend to science At best they represent a compromise between the rational and the instinctive, drawing as they can on elements of the scientific that lie snugly adjacent. Internally their promise of development lies in a frank recognition that theirs is an art; that, as in clinical medicine, their art is unworthy of the name if any possible conjunction with science is overlooked or set aside Externally their hope lies in shaping public opinion by the unswerving integrity of their practice and the truth and cogency of their precepts The basic principles should be and are self-evident They should both reflect and illuminate the democratic ideal.

Five basic principles for a merit system of public administration may be said to have reached the stage of general acceptance In a democracy they are axiomatic· (1) open competition as the method of choosing public employees, (2) selection by practical and scientific methods of the most competent personnel available from open competition, (3) equal pay for equal work, (4) a career service conditioned upon meritorious performance of work, and (5) the right of appeal from any or all personnel actions

It is unfortunate but not less than human that disagreements over the minutiae of carrying out a principle sometimes become so noisily vociferous that the still, small voice of the principle is lost in the general clamor. Mere contentiousness, however, is a minor ill. More critical distractions from the principles arise from legitimate differences of opinion and the swift crosscurrents of deadlines to meet, techniques to refine, inefficiencies to scotch, rituals to perform, red tape to unwind, irate applicants to soothe, and special privilege seekers to confront.

In the hurly-burly it is sometimes forgotten that merit systems are never ends in themselves. They exist to serve the operating agencies But merit systems should be given administrative strength equal to that of the operating programs and should have their unqualified support. Lacking this, they may in time degenerate into irresponsible and malodorous adjuncts of the spoils system. The relationship between the operating agency and the merit system is reciprocal As the one prospers, so does the other. If the one sinks into respectable somnambulism, the other is desolate also.

Some of the complexities involved in applying basic merit principles will be apparent if we examine each principle separately (1) The principle of *open competition* as a generalization is especially alluring in a democracy. When H L Mencken in his salad days translated the Declaration of Independence from sonorous Johnsonese into American, he voiced the underlying principle in language dear to the heart—"Everybody is as good as everybody else and maybe a damn sight better."

In greater or lesser contradiction of the principle, minimum qualifications of education or experience or both are often set up as obstacles to competition The rationalization is that these hurdles constitute protection against the cost and waste of failing large numbers of applicants, experience having shown that most people with 4th grade education and wheelbarrow work-history flunk out on examinations for white-collar jobs Persons of great natural endowment are, of course, less dependent upon formal education than others. Americans like to think, however, that Lincoln, the Wright brothers, and Edison are typical rather than exceptional. So-called minimum qualifications may sometimes be subordinated to the interests of special vocational or professional groups; if severe restrictions are imposed, they may become subversive of the principle of competition they were designed to facilitate. The danger is accentuated by the fact that amount rather than quality of experience is credited (Comparable data on quality of experience are almost unattainable.)

Local residence is almost completely indefensible as a hurdle to competition Yet it is widely accepted, perhaps because it expresses the yearning of the community to regard itself and its members as *better than* . By way of illustration it is rumored that every faithful citizen of Boston hopes some day to move to New York and not to like it!

Preference for special groups limits the application of open competition. Lady Bountiful scattering sunshine by her presence and gifts from her well-stocked larder is a national *idol*. Except in poetry we seldom think that "Earth gets its price for what Earth gives us." For example, recent polls show extensive public support for continuing the tradition of preference in the public service for veterans, their wives, and kin. What would be the effect on public psychology if proposed legislation read about as follows If you are not a veteran, you will have 10 points subtracted from your competitive score on whatever examination you take for a position in the public service, *if you* are a veteran but not wounded, you will have 5 points subtracted. It is not only in our efforts to befriend veterans that we fall into the illogical quicksand of something for nothing. We do so elsewhere. All of us are appalled at the "mounting death rate owing to heart disease." We are brought up short, however, when we are asked what death rate we should like to see rise if that for heart disease could be reduced

(2) *Selection of the best available* as a result of competition commends itself to the rationally minded. Most commonly the process consists of a written examination, a rating of education and experience, and an oral interview. Resulting scores are combined by differential weighting, each weight presumably determined from the relative appropriateness of the part scores to the requirements of the job. At the risk of over-simplification, we may regard the written examination as a check on the powers of thinking and the extent of appropriate knowledge, the rating of experience, a check on stability—the staying powers of the candidate; and the oral interview as a check on his histrionic abilities—how good he is as an actor.

It is in the area of the written examination that personnel administration today comes most closely into conjunction with science Full advantage should be taken, but often is not, of this fact Scientific measurement has come about within the memory of men still living. Between the early attempts of Galton and Cattell to apply to human traits measures analogous to those used in physical science and the appearance approximately two decades later of the Army Alpha lies a period of significant advances. Since that time, in schools, colleges, and in our armed forces this science has come of age, both in theory and in application. It has been expanded to include special skills and aptitudes, attitudes and interests, and is being extended to include relatively unique differential traits In merit administration the advantages this science offers in "arranging candidates in order" is recognized

Rating education and experience has been an armchair exercise. About all that is reliably measured is amount of time spent in school or at work; the quality of the experience and its effect upon the person are ignored True, those who have gone to school for a long time or have held responsible positions are, if a random group, more likely to succeed in similar circumstances than a corresponding group of the same age who have not. But extensive studies by competent observers have shown, for example, that there is a college for every level of ability in almost every State, and that variability of achievement within colleges is marked. But neither of these facts is considered in evaluating education Although the studies of work experience have not been so extensive nor perhaps so carefully made, little strain is put on the imagination in recognizing that evaluating work-experience and schooling are equally complex. What is true of the one would more than likely be *true of the other.*

That too much reliance can easily be placed upon work-experience is made clear by three commonly practiced administrative devices plus a factor that we shall call "survival" When an administrator has on his staff a difficult or inept person—but not so bad that dismissal is, under the rules of tenure, relatively simple—he watches his opportunity to "unload" that employee on someone else. The inept one moves on to his new position happily unaware of what has happened to him. And he may be similarly unloaded several times, carrying with him as he moves crabwise a record of lengthening and satisfactory experience. If he were given an unsatisfactory record, he could not be unloaded. The second administrative practice is familiarly known as "kicking upstairs." For equally good, or bad, reasons the unwanted individual is eliminated through promotion or reassignment within the organization but at a *higher* level. The third practice is "holding down" That is, an employee may be so valuable in a low position that the administration cannot afford to advance him In any event the record of the worker, as seen by an objective observer, masks at least a *part* of the truth.

As for the factor of "survival," it is among the commonest of influences contributing to unreliable rating of *work-experience*. In every organization people tend to get bored. Those most ambitious, energetic, and competent will find opportunities for advancement more readily than others. If top management by too little or poor leadership contributes to this centrifugal force, the better people rather than the weaker ones will tend to find elsewhere jobs more suitable to their talents. These vacated jobs are likely to be filled from among those who are left. Eventually persons having long, solid, C+ *experience* predominate. This is the factor of survival that over a period of years contributes to high ratings on experience for mediocre individuals and, incidentally, may staff the top jobs with incompetents.

Oral interviews could be a very useful part of the selection process Although they are extensively used for positions requiring contacts with other workers and with the public, their value is sometimes questionable. Part of the difficulty results from failure of examiners to recognize the limitations of the interview and to refine techniques that will point up its virtues. Essentially the interview should be a check on the candidate's *powers of acting*. That everyone is an actor will be immediately apparent if we reflect for a moment on what happens during every conversation We keep the lid on those things that would give offense and try to express only those that will *please or that will accomplish* our ends with the least amount of friction. When the "oral" is used as the casting director employs the tryout, we shall have begun to refine it for more *effective* use. The casting director asks the people who come before him to "register" this or "register" that. His applicants, who aim to please, respond accordingly; and the director then makes a selection from among them according to the *requirements* of the part that is to be played "The part" and "the requirements of the part."—these are the terms that need definition in any development of devices to arrange competitors in relative order of skill

To sum up, selection normally consists first of a written test. Such a test can and should be carefully prepared since a science for doing this is available. It is the best method of determining the would-be public servant's knowledge and his ability to use that knowledge The rating of training and experience is a second factor contributing to the selection process Has the applicant been through the conventions of society, both educational and social, that are thought to be prerequisites to competent performance on the job? Sole reliance on these conventions, prerequisites, or minimum requirements implies a faith in two other propositions patently false, namely, that all individuals have equal endowment and that all respond equally to experience. If these last two propositions are not true, then there is always a real probability that someone can be found who could do a competent job although lacking minimum qualifications Nevertheless, the establishment of some conventions is administratively desirable and even essential But their promulgation by fiat should be made in a state of self-awareness rather than in a state of self-delusion These conventions simply apply two questions to the competitor Is he reasonably well equipped for the position for which he is competing? Does his experience indicate dependability? Because of limitations in evaluation, in the individual, and in society, such ratings are purely judgmental today, that is, they are still unscientific Finally, the oral interview can and does contribute to the selection process. *Its purpose needs clarifying* and its techniques refining The three in optimal combination yield better results than any one alone

(3) *Equal pay for equal work* is a necessary complement to open competition and selection of the best qualified if a merit system is to be productive of an effective working spirit. Few forces are more disruptive of staff morale than inequities in pay

and rank. From the principle of just recognition for work done *stems the durable structure* of an organization, the position classification and pay plan In principle equality and fairness are axiomatic. They need no defense. But they are as difficult and elusive to apply as they are self-evident. There is as yet no science of organization, no clear dependence of pay scales on ineluctable fact Even more distressing is the apparent ease with which known elements of human psychology and close-to-earth experience can be ignored.

If keen observation and fidelity to meaning of facts can be called psychology, the novelists are often to be classed among the best psychologists As Charlotte Bronte wrote in the "Editor's Preface to the New Edition of Wuthering Heights":

> The writer who possesses the creative gift owns something of which he is not always master—something that, at times, strangely wills and works for itself He may lay down rules and devise principles, and to rules and principles it will perhaps for years lie in subjection, and then, haply without any warning of revolt, there comes a time when it will no longer consent to "harrow the valleys, or be bound with a band in the furrow" When, refusing absolutely to make ropes out of sea-sand any longer, it sets to work on statue-hewing. . Be the work grim or glorious, dread or divine, you have little choice left but quiescent adoption. As for you—the nominal artist—your share in it has been to work passively under dictates you neither delivered nor could question. . . If the result be attractive, the World will praise you, who little deserve praise, if it be repulsive, *the same World will blame you, who almost as little deserve blame.*

On first reading one gets the impression that here is, perhaps, the best definition of genius he has seen But closer reflection, stimulated by the shift in point of view from the impersonal "he" to the very personal "you," suggests that Charlotte's description is not of genius at all but of everyone who is human The ordinary or garden variety of classification and pay plan under a merit system leaves much undone that might with a little effort mold such plans closer to the heart's desire. The typical pyramidal concept leaves out of account much of the universal human urge for elbow room—freer play for talent: At the top administration, next supervision, then professions and specialized vocations, and finally the workers. Even at the top it is quite possible to search dozens of plans

and not find one that would make provisions elastic enough to attract a man like William James—physician, psychologist, philosopher, man of letters; or Benjamin Franklin—artisan, scientist, inventor, diplomat, statesman

Inelasticity, however, is not the only weakness Of greater jeopardy to the principle of fairness, but perhaps *more easily* remedied, is the failure of many operating administrators to see the manifold advantages a classification and pay plan can contribute to good administration. Too often they regard the structure as a hampering rather than a facilitating device; by circumventing the requirements they put themselves in the position of the sick man who brushes aside his physician and clings to his amulet Merit systems must contribute more than they do now to informing and instructing the operating agencies—administrators and workers alike.

(4) Fourth among the major principles of merit administration is the requirement that the system provide for a *career service based on merit.* Basic to this principle is tenure of office for those who have been selected as the best available and who have qualified for permanent appointment by having served successfully a working-test or probationary period Once they have passed muster it is in the interest of the public service that they be given reasonable assurance of holding their jobs. With this assurance they are free from fear, and can direct their energies to increased job efficiency. Bernard Shaw, however, cautions those who start at the bottom and, hopefully, climb the proverbial ladder rung by rung. You don't learn even to hold your own, he says, "by standing on guard but by attacking and getting well hammered yourself"

A policy of promotion-from-within is another element of a career service. It is carried to extremes when it is so nearly like the human circulatory system that elaborate preparations and intravenous injections are necessary for the purpose of introducing a little new blood. Very often new blood brings new life, as many a soldier can testify "Permanent tenure" is not an unmixed blessing It has even been called a necessary evil. Those who accept it as a principle of administration and those who profit from it as employees should understand its hazards

and the limitations on its effectiveness As part of a career service, some method of evaluating the worth of those with the prospect of promotion is necessary Recourse here to the basic principles of selection is indicated—written promotional examinations with a broad base of competition, plus service or efficiency ratings as partial determiners of "the best available."

Service ratings leave much to be desired. They have inherent limitations. They put a premium on conventionality and often on the mediocrity that never swerves from the beaten path They do not foster understanding of the brilliant employee whose contribution this week is more than noteworthy but whose unexpected absence the following week leaves the administrator gnashing his teeth Service ratings do, however, bring out into the relative open, so that impartial observers can see them, the supervisory judgments that would otherwise be made behind closed doors and acted upon arbitrarily in star chambers Moreover, the rating process can be refined If employees share in it and understand its advantages as well as its limitations, it can serve an exceedingly useful though limited function

(5) The last of the five basic principles is the *right of appeal.* This right rests on the theory that the individual under merit administration has basic rights that cannot be ignored and that the administrator has responsibilities that he must discharge Appeals usually lie when an applicant has been refused the privilege of standing an examination; or if he feels that he has been unfairly rated on one or more parts of the selection process, or if he challenges the fact that he was not certified or that his name was improperly deleted from a register Appeals frequently lie for layoff, demotion, suspension, and almost always for summary dismissal

Many defensible differences of opinion exist over particulars in every system of appeals. the actions covered, the nature and calibre of the appeals body and the manner of appointment, the question of informing the worker of his rights, and the question of the administrator's responsibility once an appeal has resulted in a finding But there is little if any disagreement among merit or civil service systems that the right of appeal is cardinal.

In this article merit systems have been criticized—not that they may be destroyed but that they may be improved The only alternative to selection by merit is selection by personal whim In that direction lies chaos Thus, society is left with the task of making its "merit system" more workable, more reliable, more internally consistent—in short, more nearly scientific. Enough experience with merit systems is now available that some of the malfunctionings are apparent Knowledge is available that could be used to correct these malfunctionings. Improvement of merit systems is contingent upon their scientific development, their acceptance is dependent upon a greater public awareness of what merit systems are, their perfection waits upon a demand that they be what they could be

In *The Faith That Heals*, Sir William Osler, noblest physician of his day, says, "Nothing in life is more wonderful than faith—the one great moving force which we can neither weigh in the balance nor test in the crucible" Through its faith in democracy this nation is philosophically committed to the merit principle Imperfections are transitory. The "substance of things hoped for" is scientific selection of the best qualified public servants

---

HOW TEACHERS CAN IMPROVE THEIR TESTS[1]

MAX D ENGELHART[2]
Chicago City Junior College

THE chief function of a teacher is that of directing and motivating pupils toward the attainment of desirable educational objectives. In the performance of this function testing can play a very important part When objectives are adequately defined and tests devised which are valid with respect to them, the extent to which objectives are being attained can be measured. Furthermore, such tests define the objectives for the pupils and motivate the pupils toward them When the results of such tests are adequately analyzed and interpreted, the teacher obtains a means of better orienting instruction and the pupils secure motivation through knowledge of progress

Construction of exercises and analysis of the data resulting from their use can make objectives more definite and more meaningful to the teacher The creation of novel exercises and the analysis of data pertaining to them may widen the scope of objectives recognized and ultimately realized in instruction

Instructional objectives are most usefully defined in terms of observable behavior. Each specific objective should be an answer to the question "What should the pupils be able to do as a result of instruction?" Instruction which produces the abilities to do certain things, should concomitantly develop the attitudes, interests, and ideals which motivate their doing. Instead of the general and intangible objectives "good citizenship," "appreciation of good literature," and "scientific method," specific objectives formulated in terms of observable behavior may include. "Presenting arguments in support of the elimination of the general property tax based on factual evidence critically analyzed and evaluated," "selecting a short story for leisure reading on the basis of the following criteria. . ," "rejecting a conclusion which goes beyond the data"

Many instructional objectives must be concerned with factual information, or knowledge In the selection of facts consideration should always be given to the contribution such knowledge can make to the types of behavior illustrated in the preceding paragraph The thinking necessary for adequate performance of activities recognized as the really worth-while objectives of instruction must be based on knowledge. Information which makes no recognizable contribution to such thinking, or to the further learning which may contribute to such thinking, is not worth teaching. It is also not worth testing

While teachers usually contend that their objectives are not restricted to the memorization of miscellaneous, unrelated, and often trivial information, the tests used by most teachers are convincing evidence that their actual objectives are thus restricted When objective tests are made by teachers the exercises are most often of the true-false or multiple-answer types and the content of these exercises is wholly based upon information given in the text When essay tests are constructed and used, the questions are factual in character and are scored only for the facts remembered.

Tests should be designed which measure knowledge of facts True-false or multiple-answer exercises can be efficient means of measuring such knowledge. The possibility of representative sampling of pupil knowledge is one of the advantages of objective testing In writing such exercises the teacher should be critical in the selection of the facts to be tested One useful criterion in the selection of facts is the degree of relevance of the fact to some important general concept or principle which can be applied in the solution of some problem involving reflective thought. Consideration should be given to whether or not knowledge of the specific fact contributes to further learning In many fields progress in learning is contingent upon the synthesis in the mind of the pupil of an ever growing body of factual knowledge In thinking about an important contemporary social problem the pupil may require a knowledge of numerous historical facts relevant to the trend which has created the problem Series of factual objective exercises may be useful in determining the extent to which such knowledge has been attained.

In writing true-false exercises certain precautions should be observed Broad generalizations should usually be avoided. Such words as "always," "never," "none," "only," "all," and "every" are obvious clues to the falsity of certain statements when the pupil can readily think of exceptions On the other hand, the statement "all echinoderms live in salt water" represents a difficult true-false item to one whose biological knowledge is not extensive. In multiple-answer exercises the incorrect completions should not be too obviously incorrect. Each completion should be plausible It is frequently effective to write multiple-answer exercises in which a "best" rather than a "correct" answer is called for. Such exercises tend to accomplish more than the measurement of memorized information. All, or most, of the completions may be "correct"; the pupil must judge which completion is the "best." In a recent social science comprehensive examination the directions for one series of multiple-answer exercises asked the student to identify the alternative suggesting the "*most* significant relationship" between the two things mentioned. The first exercise in this series is given below

1. Minor party—social and economic reform· (A. Minor parties are usually characterized by radical platforms; B minor parties seldom win an election, C reforms agitated by minor parties are sometimes adopted and enacted into law by the major parties, D unsuccessful minor parties ultimately pass out of existence, E. the major parties have the advantage in organization, funds, and prestige and, hence, are more successful in promoting reforms)

While most of the answers are "correct," alternative "C" represents the most significant alternative In a recent biological science comprehensive examination each exercise began with a statement to be explained. More than one of the alternatives were "correct," but only one was accepted as the "explanation." The following exercise is an example:

---

[1] Reprinted by permission of the *Chicago Schools Journal*

[2] On leave as a member of the Examinations Staff of the United States Armed Forces Institute

Statement In embryological development certain structures develop in a manner analogous to the conversion of a passenger boat into an aircraft carrier (A The morula changes into a blastula which changes into a gastrula, B part of the placenta is formed from the wall of the uterus of the mother, C the mesoderm gives rise to the heart, blood vessels, blood cells, lymphatics, kidneys, and certain other organs and structures, D in the higher vertebrates, the middle ear cavity, the eustachian tube, the thymus gland, and the parathyroid glands develop from gill slits or arches, E the ectoderm gives rise to the epidermis and to the lining of the mouth and anal apertures)

The answer "D" is the alternative which best explains the statement. This example illustrates another very important matter If the teacher desires to measure how well pupils can handle questions involving thought rather than memory, an exercise must constitute a novel problem. It is improbable that the students were taught the particular analogous relationship implied in the above exercise. If they were taught this analogy, the exercise will measure only the extent to which the analogy is remembered If, however, they were merely taught the facts represented by alternative "D" then the relating of this alternative to the introductory statement becomes an act of thought transcending the utilization of memory. The examples just given also illustrate the fact that exercises which are simple in form can be effective in the measurement of behavior involving thinking. It is not necessary to construct exercises involving complex directions in order to obtain such measurement. The important thing is that the content of the exercise should be to some extent novel to the pupils

Discriminative Thinking

One means of securing the novelty referred to in the preceding paragraph is to bring together in a series of objective items numerous facts whose classification in certain ways will involve the ability of the pupil to do discriminative thinking and to synthesize his knowledge. Let us suppose that pupils in physics have been taught certain facts pertaining to sound and, at a later date, certain facts pertaining to light Let us suppose further that the teacher has not stressed the similarities and differences of sound and light phenomena The good

teacher would probably stress these things, but for the sake of our illustration, let us suppose that this is to be postponed until the test has been given If this has been the case then the following series of *exercises should involve discriminative* thinking and, as this thinking is taking place, require a synthesizing of knowledge by the pupil·

On the line preceding each of the following items, write the letter

A if the item is true of sound
B if the item is true of light
C if the item is true of both
— Its velocity in water is greater than in air.
— It can be reflected.
— It can travel through a vacuum.
— It can be refracted
— Etc.

A number of examples of this very useful general type of exercise are given below from various fields In each case only a few of the items are listed

In each situation below, an individual or a group of individuals is seeking protection or assistance On the line preceding each of the following items, write the letter

A if the Federal government is responsible
B if the state government is responsible
C if both governments are responsible
D if neither government is responsible
— The New York Life Insurance Company wishes to open up agencies and sell insurance in Oregon.
— Mr. Jones receives in payment $1,000 in bills which he presently learns are counterfeit
— A Chicago visitor from Fort Wayne. Indiana, suffers severe injuries when his car is wrecked because of defective pavement on 76th Street
— Convicts escaped from Joliet Penitentiary arrive in Des Moines, Iowa, hold up a bank, and are seized and held by the local authorities.
— Etc

On the line preceding each of the following items, write the letter
A if the sentence is fragmentary
B if the sentence contains a comma fault
C if the sentence contains a dangling modifier

D if the sentence exemplifies lack of parallelism or faulty parallelism
E if the sentence is correct
— One day I feel as though I could lick the world, the next day I feel like a swatted fly
— Upon returning from the store, my homework requires my attention.
— As we walked along the hall, where a large photograph of Roosevelt hung.
— All of the boys being gone, most of the manual labor was done by the girls
— Bemoaning her stupid lot, the army and navy claimed all of her friends
— A freshman learns to study with regularity, to play with enthusiasm, and co-operation
— Etc.

On the line preceding each of the following items, write the letter
A if the statement is true of the *Lycidas*
B if the statement is true of the *Adonais*
C *if the statement is true of both works*
D if the statement is true of neither work
— The poem was written in memory of a dead friend
— In form as well as content the poem displays a deep personal grief
— The poet ornaments his verse with many classical allusions.
— Etc.

The following type of exercise can be used in a variety of school subjects and is effective in measuring how students can make *comparisons:*

On the line preceding each of the following paired items write the letter

A if the item at the left of the page is of greater magnitude than the item at the right
B if the item at the right of the page is of greater magnitude than the item at the left
C If the two items are of equal magnitude

—	Amount of glucose in blood.. entering the liver 4 hours after a meal is eaten	Amount of glucose in the blood leaving the liver 4 hours after a meal is eaten.
—	Amount of absorption of foods. ... by stomach.	Amount of absorption of foods by intestine.
—	Percentage of urea in blood.... entering the liver	Percentage of urea in blood leaving the liver.

—	Amount of heat produced in. . the body.	Amount of heat retained in the body

— Etc.

The *same form needs only slight adaptation to be useful in* writing chronology exercises in history

On the line preceding each of the following paired items write the letter
A if the event in Column I occurred before the event in Column II
B if the event in Column II occurred before the event in Column I
C If the events occurred at approximately the same time (within about a year of each other)

	COLUMN I	COLUMN II
—	Clayton Antitrust Act	Sherman Antitrust Act
—	Alabama Claims Case	Venezuelan Arbitration
—	Dred Scott Decision	Fugitive Slave Act
—	"Wigwam Convention"	Fort Sumter fired upon

— Etc.

It should be mentioned that there should be some relationship between the paired events significant enough to warrant their being paired Care should be exercised when writing the "C" items to give events that occurred simultaneously, or very nearly simultaneously Note the qualifying remark with respect to category "C" in the directions stated above

Exercises of similar form are useful in measuring the way in which pupils handle correlated or cause and effect relationships In writing such items where the relationship is definitely cause and effect, the cause should be given first.

On the line preceding each of the following paired items write the letter

A if increase in one of the things referred to is usually accompanied by increase in the other
B if increase in one of the things referred to is usually accompanied by decrease in the other
C if one of the things referred to tends to remain the same when the other increases or decreases
— Amount of carbonates dissolved in the water of a river Number of clams in the river.
— Temperature of the environment of a bird or mammal. Body temperature of the bird or mammal.

__ Amount of dissolved salt (sodium chloride) in a given body of water Number of amphibia in the water
__ Extent to which men make changes in an area Rate at which the area tends toward balanced equilibrium
__ Etc.

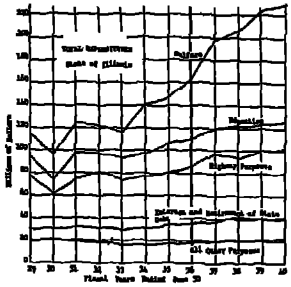
On the line preceding each of the following paired items, write the letter

A if significant increase in one of the things mentioned has usually been accompanied by significant increase in the other
B if significant increase in one of the things mentioned has usually been accompanied by significant decrease in the other
C if one of the things does not tend to change significantly when significant change takes place in the other

__ Dollar diplomacy *Confidence in the United States on the part* of the South American Nations.
__ *Efforts to eliminate the poll tax. Oratory on state rights by* Southern senators
__ *Efforts to liberalize the Supreme Court. Expressions of con*servatism with respect to the issue on the part of the general public.
__ Acceptance of the doctrine of checks and balances Lobbying.
__ Etc.

Intelligent Reading

The ability to read intelligently in a field is an important general objective and more *efforts should be made to measure* how well pupils can perform this activity This ability is a very important factor in further learning within a field and in dealing with practical problems when formal school learning has terminated. The following exercise illustrates one means of measuring such ability. *In this example, the material to be* read is a graph. Such exercises need not be so restricted One may use paragraphs drawn from texts other than those studied by the pupils and even from advanced texts if the teacher wishes to challenge the pupils. When paragraphs are selected *they should be self-contained in that the topic is* treated fairly completely. It is frequently effective to present paragraphs which give scientific data and some of the statements listed may be inferences which may justifiably be derived from the data. Others of the statements may be only partially justified or may be irrelevant. *Some of the items of the following exercise represent predictions which may not be justified:*



On the line preceding each statement write the letter

A if the information given in the chart is sufficient for a judgment that the statement is definitely true
B if the information given in the chart is sufficient only to indicate that the statement is probably true
C if the information given in the chart is sufficient for a judgment that the statement is definitely false
D if the information given in the chart is sufficient for a judgment *that the statement is probably false*
E if the information given in the chart is not sufficient to indicate any degree of truth or falsity in the statement

__ Less money was spent in 1930 than in 1929 for welfare and education.
__ In 1931 and 1932 the expenditure of money for highway purposes was evidently considered a means of combatting the Depression.
__ In 1940 a much greater proportion of the total expenditures was for welfare than in 1942
__ Had our country not entered the war in 1941, expenditures for welfare in 1942 would have been greater than in 1940.
__ The increasing amount of money spent by the State for all purposes between 1929 and 1940 must have come largely from taxes or from Federal grants-in-aid rather than from borrowing.
__ Etc.

Such exercises can be *scored in the usual way to obtain a* total score which shows how well the students agree with the key set by the teacher, or the "expert thinker" in the field. Such a total score may answer the questions "To what extent does the pupil read data accurately?" and "How well does the student *recognize valid generalizations drawn from data?*" It is possible, however, to obtain other scores which reveal certain characteristics of pupil thinking. For example, a count of the number of "B" items marked "A," and "D" items marked "C," and "E" items marked with something other than "E" is indicative *of the extent to which the pupil tends to go beyond the* data A relatively low score would be indicative of relatively greater maturity in thinking with data. Such analysis is one of the characteristics of the handling of the test data in the Eight Year Study of the Progressive Education Association and in the Cooperative Study *in General Education of the* American Council on Education. The following statement appears in a recent report of the Board of Examinations of The University of Chicago.

> The modern tendency is to contruct the items so that the wrong responses are wrong in a specific way—for example, a definition which is too broad, or a definition which is inadequate. When alternative answers are so chosen, a regular pattern of incorrect responses is established which upon analysis yields much more information about *the students' mental habits than did incorrect responses constructed* by the older methods. By means of this pattern type of analysis, it is possible to determine whether students' errors in reading and interpreting data consist in saying that certain things are in the data which in fact are not, or in saying that certain things are not in the data which in fact are. It may be found that students are able to select statements which agree with the data presented but have difficulty with statements which disagree with the data presented.

It is *difficult to write exercises of the types described* Such exercises must be written and used, however, if pupils are to seek worth-while objectives and if the degree of attainment of worth-while objectives is to be measured Where measurement is restricted, objectives are also restricted.

When a teacher has written such exercises it is essential to secure careful evaluation by other teachers who have the same general objectives. The best evaluating is done when the evaluating teacher does not have access to the key, but answers the items herself Comparison of several such evaluations is valuable in the rejection of bad items and in the revision of others. For example, in a biological science examination the students were asked to *mark certain items true or false* on the basis of the "principles of inheritance" The following item was accepted as false by several biological science instructors: "Boys tend to resemble the father, while girls tend to resemble the mother" However, one instructor pointed out that in a certain *important respect* boys *always resemble their fathers* and girls always resemble their mothers In the directions preceding the items the following qualifying phrase was added to take care of the situation: "excluding primary and secondary sex characteristics"

The *preceding* paragraphs have dealt *exclusively with objective exercises* In any balanced program of testing some essay exercises should be included. In writing such exercises "fact" questions should be avoided. Objective exercises can test knowledge of facts more efficiently and representatively *than essay questions. Essay exercises should represent novel* problematic situations For example, the following essay exercise appeared in a recent physical science comprehensive examination:

> In 1492, Christopher Columbus began his voyage of discovery by sailing *southwest to the Canary Islands which are near the coast of* Africa and in latitude 28° N. He then continued his voyage by sailing westward in that part of the Atlantic Ocean between the equator and 30° N On his return trip to Spain, early in 1493, he first sailed northeast until he was somewhat more than 30° N and then sailed east to Spain. On the basis of information given in your physical science text, explain why Columbus sailed as described above.

Several blank lines followed this exercise in the test booklet. *Nothing is said about Columbus in the physical science text,* but information on the belts of the winds is given which could be applied by the student in responding to the question Columbus took advantage of the northeast trades in his voyage to the New World and of the prevailing westerlies on his return *to Spain.*

Essays have been based on selections of quoted material and on cartoons. In the field of English composition it effective to have essays based on notes presented for reading at the time of the examination, or prior to the time of examination.

Information is needed for the writing of a correct response to essay exercises which are thought questions or novel problems, but the quality of the response will also depend upon the extent to which the student critically analyzes the situation and thoughtfully organizes his information in an effort to meet it. The scorers should be sensitive to more than the correctness of the information presented by the student. Their ratings should be based not only on the correctness of the facts, but also upon evidences of superior selection, evaluation, and organization of the facts presented. Comparison of student responses with a scale of responses to the same or to a similar question may be found to be an effective procedure. Another possibility is the use of directions for scoring in which such characteristics as organization and originality are defined and illustrated, and suggestions are made with respect to the weights to be given to each such characteristic.

Analysis of Test Scores

After a test has been given and scored the test data should be subjected to analysis. Analysis is essential if the teacher is to know the extent to which objectives are being attained. One type of analysis has been referred to in the paragraph following the exercise based on a graph. It is also very desirable to determine the per cent of correct response to each exercise for the group taking the test. A low per cent of correct response may indicate the need of further instruction. In some cases low per cents of correct response may warrant the rejection of such items for use in testing subsequent classes, or the omission of such subject matter as inherently too difficult. The analysis may be extended further than merely determining per cents of correct response. One can, with a little labor, determine how well each exercise correlates with whatever is measured by the test as a whole.

One way to do this type of analysis is to separate the papers into two groups. The "upper group" contains all papers above

the median score of the test as a whole while the "lower group" contains all papers with total scores below the median total test score. Taking one test paper at a time and opposite the numbers of the exercises on a tally sheet, the teacher tallies for each exercise correctly answered. For example, if the first paper has correct answers for exercises 1, 2, 5, 7, and so on, the teacher makes a tally mark after these numbers on the tally form. Another form is similarly prepared for the "lower group." The per cents of correct response for each of the groups are then computed. (Samples of 100 papers in each group avoid the necessity for such a conversion.) Since there are equal numbers of papers in the upper and lower groups corresponding per cents may be averaged to obtain the per cents of correct response for the entire group taking the test. The per cents for the upper and lower groups are used in reading the correlation coefficient from the abac shown on page 122.

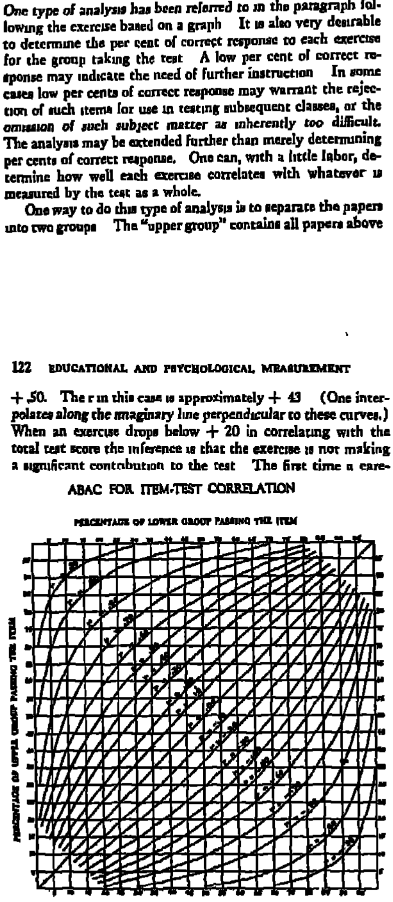
For example, suppose that 65 per cent of the upper group answer exercise 17 correctly while only 25 per cent of the lower group do so. The correlation between success or failure on the item and the total score on the test is +.60. Such an item makes a significant contribution to whatever is measured by the test as a whole. This correlation can be seen in the following table which need not be constructed for each item, but which is useful in explaining the above.

	FAILURE ON ITEM	SUCCESS ON ITEM
ABOVE MEDIAN	35	65
BELOW MEDIAN	75	25

r=+.60

Take the per cents 80 and 55 as another example. Here the intersect does not fall on one of the lines in the chart. One must interpolate between the lines labeled r=+.40 and r=+.50. The r in this case is approximately +.43. (One interpolates along the imaginary line perpendicular to these curves.) When an exercise drops below +.20 in correlating with the total test score the inference is that the exercise is not making a significant contribution to the test. The first time a carefully constructed test is given it is not unusual to find that one-fourth or one-third of the exercises drop below this value. Certain exercises may yield negative correlations. For example, suppose that only 25 per cent of the upper group answered exercise 48 correctly while 65 per cent of the lower group responded correctly. The correlation is −.60 and this negative relationship is illustrated by the following table:

ABAC FOR ITEM-TEST CORRELATION



	FAILURE ON ITEM	SUCCESS ON ITEM
ABOVE MEDIAN	75	25
BELOW MEDIAN	35	65

r=−.60

Low or negative correlations most often indicate that an exercise is bad. The key may be wrong, the exercise may be ambiguously stated, or the exercise may be much too easy or much too difficult. Analysis of responses to other alternatives can determine a better key when the exercises are of the type requiring judgments; for example, the type of exercise illustrated with reference to the reading of the graph of state expenditures. In some cases a low or negative correlation does not necessarily mean that the exercise is a poor one. It is possible that the exercise is measuring some trait which is worth while in itself, but which is not related to whatever is measured by the test as a whole. In every case the teacher should study the exercise in relation to the total per cent of correct response and the correlation coefficient, and should formulate judgments with respect to the merit of the exercise and the degree of attainment revealed by the data.

The type of analysis just described applies to objective exercises. No definite procedure can be suggested for essay exercises. When the scores on a given exercise range over several points, ordinary Pearson product moment coefficients can be calculated between the score on the exercise and the score on the test. It would seem much more useful, however, to analyze essay responses in terms of some classification with respect to various types of merit or of limitations.

The discussion in the preceding paragraphs may seem to be carrying the "improvement of tests" a bit too far for the classroom teacher. It may seem that an inordinate amount of work and cerebration is involved If, however, teaching is as much worth doing as the producing of manufactured articles or materials where the processes are controlled from stage to stage, then the labor involved is more than justified The teacher who is willing to do these things will find that the "improvement of tests" is not merely that. It is also the improvement of teaching.

PREDICTION OF COLLEGE SUCCESS BY MEANS OF THURSTONE'S PRIMARY ABILITIES TESTS[1]

CHARLES H. GOODMAN
The Pennsylvania State College

In his 1940 monograph on factor analysis, Wolfle (18) states: "Few attempts have been made to put the results of factor analysis to practical use Most of these few have dealt with intelligence testing . Some studies, Chein (5) and Schneck (9), have indicated that achievement in individual courses can sometimes be predicted as well [as], or better, by special tests or by specialized subtests than by the total score of a general intelligence test. Which tests will best predict grades in individual courses is an empirical problem As a start toward answering that problem, Thurstone has prepared a battery of sixteen tests, giving perception, number, verbal, space, memory, induction, and deduction scores (15). He warns (16) that this battery is still in the research stage and is not ready for routine use. It has been used by Bernreuter (3) and Stalnaker (11), but no data have been published to indicate how well it answers the administrator's wish for better methods of predicting scholastic achievement."

This paper is a report upon the studies conducted at The Pennsylvania State College specifically related to the possibilities of using the Thurstone *Primary Abilities Tests* for the purpose of predicting scholastic achievement, it provides, in some measure, a partial answer to the statement made by Wolfle (18) that there is a need for data to indicate how well these tests answer the administrator's wish for better methods of predicting scholastic achievement

---

[1] The writer wishes to acknowledge his thanks and appreciation to Marianne Hassemar, Virginia Dickey Tredick, Isabella Waddell White, and Fred J. Ball for their permission to report the findings of their studies related to this paper

Experiments and Findings

Bernreuter and Goodman (4) conducted a study in 1939 to determine how well they could predict the success of freshmen engineers at The Pennsylvania State College by means of the Thurstone *Primary Mental Abilities Tests*. One hundred seventy freshmen engineers served as their subjects. Bernreuter and Goodman then obtained correlations for the four college courses of Chemistry, Drawing, English Composition, Mathematics, and semester point average[a] with the *Primary Abilities Tests*. They then calculated multiple correlations, using various combinations of the abilities, with semester point average, Chemistry, English Composition, and Mathematics. The results are shown in Table 1.

TABLE 1

Correlations and Multiple Correlations of the Primary Abilities with College Courses in Engineering

Ability	Semester point average	Chemistry	Drawing	English Composition	Mathematics
P	+.04	+.07	+.00	+.05	+.04
N	+.32	+.27	+.01	+.26	+.27
V	+.33	+.32	+.01	+.44	+.16
S	+.23	+.19	+.11	+.11	+.25
M	+.10	+.04	+.11	+.23	-.05
I	+.34	+.23	+.18	+.21	+.29
D	+.38	+.41	+.13	+.21	+.44
NVSID	+.51	...	...	...	...
NVID	...	+.49	...	...	...
NVMID	...	...	...	+.49	
NSID	...	...	...	...	+.49

The highest single correlation obtained was +.44 for reasoning with Mathematics and +.44 for verbal ability with English Composition By using a combination of the number, space, induction, and reasoning abilities, and calculating a multiple correlation with Mathematics, they were able to obtain a multiple correlation of +.49. In the case of English Composition, by combining the abilities of number, verbal, memory, induction, and reasoning, they obtained a multiple correlation of +.49. In both cases the increase in the size of the correlation is slightly more than that obtained using the single abilities of reasoning with Mathematics, +.44, and verbal ability with English Composition, +.44 The best multiple correlation obtained, +.51, was with first-semester point average, using the abilities of number, verbal, space, induction, and reasoning. A more detailed report of this work has been published in an earlier paper by Bernreuter and Goodman (4)

During 1940 Ball (1) administered the Thurstone *Primary Mental Abilities Tests* to a group of 147 female freshmen and 159 male freshmen attending the Liberal Arts School of The Pennsylvania State College He then correlated their test scores for each of the seven primary abilities with the college courses of Art, Botany, English Composition, French, History, Mathematics, Physical Science, Political Science, Zoology, and with their first-semester point average. Ball's findings are shown in Table 2

TABLE 2

Correlations of Thurstone's Abilities with First Semester Point Average and Grades in Nine Liberal Arts College Courses

	P	N	V	S	M	I	D
Semester point average	.15	.24	.35	.04	.28	.28	.27
Art	.14	.02	.24	.11	.18	.16	.11
Botany	.13	.12	.28	.17	.18	.28	.23
English Comp.	.07	.14	.40	.09	.26	.23	.15
French	.08	.12	.27	-.10	.19	.02	.01
History	.13	.14	.36	-.12	.29	.12	.22
Mathematics	-.03	.41	.20	-.10	.16	.22	.33
Phy Sci.	.14	.20	.24	.06	.18	.26	.31
Pol. Sci.	.00	.15	.37	-.02	.32	.31	.27
Zoology	.14	.23	.28	-.04	.34	.24	.11

It will be seen from Table 2 that the range of correlations found by Ball is from -.12 to +.41. The highest correlation obtained was +.41 for number ability with Mathematics The next highest correlation, +.40, was that of verbal ability with English Composition. Verbal ability correlates more highly and positively with each of the courses than does any of the other abilities. The highest correlation Ball obtained for any of the abilities with Semester Point Average was +.35 with verbal ability In a further effort to determine whether he could increase his predictive possibilities with the Thurstone tests, Ball computed a multiple correlation of the memory, number, verbal, induction, and reasoning abilities, using as his criterion semester point average The multiple correlation was + 46. Ball's results show that optimum weights yielded only a slight increase in correlation over that obtained for the single factor of verbal ability, which correlated + 35 with semester point average

---

[a] Semester point average is obtained by dividing the number of credits earned by a student into the total number of grade points earned. While the shortcomings of semester point average as a criterion for college success are fully recognized, it appears to be the best possible one available at The Pennsylvania State College.

Hessemer (8), in a study similar to Ball's, attempted to determine the predictive possibilities of the Thurstone tests in the School of Chemistry and Physics at The Pennsylvania State College. Upon administering the *Abilities Tests* in 1942 to 147 freshmen students, she correlated their test scores on the primary abilities with their first-semester point average and with the course of Inorganic Chemistry which each of her subjects had taken Inorganic Chemistry at The Pennsylvania State College is the first course in chemistry taken by freshmen students It is both intensive and extensive in scope, as can be seen from the following description in the College catalogue·

> Inorganic Chemistry (5).—The nonmetallic elements, fundamental principles of the science are studied in connection with the descriptive chemistry of nonmetallic elements and their compounds, prepares for future study of the science. Lecture 2 hours, recitation 2 hours, practicum 3 hours

The results obtained by Hessemer are presented in Table 3 and show, as did Ball's results, verbal ability to be the best single predictor of semester point average. The best positive correlation, +.18, was that of reasoning with Inorganic Chemistry This correlation is slightly larger than four times the size of its probable error. Interestingly enough, the largest correlation found was negative, – 25, for space with Inorganic Chemistry. When each of the primary abilities was correlated with semester point average, the highest correlation obtained was that of + 44 with verbal ability, while reasoning correlated next highest, + 40

In 1941 Tredick (13) conducted a rather thorough study of the predictive possibilities of the Thurstone *Primary Mental Abilities Tests* and a battery of vocational guidance tests. Her guidance test battery consisted of the *Otis*, the *Pressey*, the *Minnesota Paper Form Board*, the *Minnesota Clerical Test*, the *Meier-Seashore Art Judgment Test*, the *Minnesota Assembly*, and the *Minnesota Spatial Relations Test*. Her subjects consisted of 113 freshmen women in the Department of Home Economics at The Pennsylvania State College. After administering these tests she correlated the seven abilities with the grades made by her subjects on the five college courses of Art, Chemistry, English Composition, Home Economics 101, Home Economics 109, and first-semester Point Average. Her findings are shown in Table 4.

TABLE 3

The Primary Mental Abilities Correlated with Semester Point Average in Chemistry School and Inorganic Chemistry

	Semester point average	Inorganic Chemistry
P	–.008	–.22
N	+.23	+.09
V	+ 44	+ 13
S	–.08	–.25
M	+.36	+.05
I	+.23	–.02
D	+ 40	+ 18

It will be observed from Table 4 that the range of correlations is from – 02 for memory correlated with Art, to +.55 for verbal ability correlated with English Composition. It will be noted that Tredick's results, like Ball's and Hessemer's, showed verbal ability correlated more highly with each of the courses than did any of the other abilities The highest any of these abilities correlated with Semester Point Average was +.51 and again it was verbal ability.

TABLE 4

Correlations of the Primary Mental Abilities with College Grades and Semester Point Average in Home Economics

	Art 76	Chemistry	English Composition	Home Economics 101	Home Economics 109	Semester point average
P	.15	.20	19	.31	18	.38
N	.11	.46	.22	.20	.33	.41
V	.24	.28	.55	.50	.37	.51
S	.25	.23	10	.22	16	.28
M	–.02	.25	.08	12	15	.20
I	.26	.37	19	.35	.26	.40
D	.21	.43	.21	.24	.33	.42

Tredick then calculated the correlations of the battery of vocational guidance tests taken by her subjects with the same five college courses which were used to obtain the correlations with the Thurstone tests, and semester point average. The findings are shown in Table 5 and afford some interesting comparisons with those obtained with the Thurstone *Abilities Tests*

TABLE 5

Correlations of the Vocational Guidance Battery with College Grades and Semester Point Average in Home Economics

	Art 76	Chemistry	English Composition	Home Economics 101	Home Economics 109	Semester point average
Paper Form Board ..	.24	.24	13	.34	17	.31
Pressey	.20	.38	48	41	43	.51
Otis ..	17	.37	.54	.48	41	.53
Name Checking	.07	.26	.07	.08	17	.23
Number Checking .	.08	.31	.27	.26	.26	.36
Meier-Seashore Art Judgment	.20	.03	.29	.34	1"	.23
Minnesota Assembly	17	.16	–.01	.06	.01	11
Spatial Relations .	.20	.23	.02	.27	.09	.23

The highest correlation obtained for a single test of the vocational guidance battery with college grades was that of the *Otis* (*Higher Examination, Form A*), which correlated +.54 with English Composition. The next highest correlations obtained were for the *Otis* with Home Economics 101, + 48, and the *Pressey* with English Composition, +.48. It will be noted that both of the group tests for mental ability, the *Otis* and *Pressey*, correlate higher with these college courses than do any of the other tests of the vocational guidance battery. Furthermore, the *Otis* correlates more highly with Semester Average, +.53, *than do any of the other tests of the vocational guidance* battery. The *Pressey* correlates only slightly lower than the *Otis* with Semester Average, +.51. Comparing Table 4 with Table 5, it will be seen that the verbal factor is the only ability that appears to correlate as well as the *Otis* or the *Pressey Tests* with the various courses The *Otis* correlates slightly higher, +.53, than does verbal ability, +.51, with semester point average. However, the correlation with semester point average of the *Pressey* is the same, +.51, as that for verbal ability

In an effort to determine the relationship between her subjects' scores on the *Primary Abilities Tests* and their scores on the vocational guidance battery, Tredick calculated the correlations for each of the tests Her results are shown in Table 6.

TABLE 6

Correlations of the Thurstone Primary Mental Abilities with A Vocational Guidance Test Battery

	P	N	V	S	M	I	D
Pressey , ...	.48	.38	.76	.40	.27	.52	.61
Otis .	.53	.33	.68	.40	.29	.60	.61
Revised Minnesota Paper Formboard	.39	.21	.24	.37	.06	.48	.45
Minnesota Assembly ..	.23	–.12	11	.34	.07	.26	.30
Spatial Relations , ...	.55	.15	16	.49	.06	.47	.33
Number Checking ..	.51	.59	.06	.36	.20	.28	.28
Name Checking	.57	.38	.40	.41	.24	.44	.46
Meier-Seashore Art Judgment	.33	.11	.39	.20	.18	.23	15

The data of Table 6 offer some indications why the correlations of Thurstone's verbal ability with the five college courses are so similar in size to those correlations obtained for the *Pressey* and *Otis* with the same college courses. Of the seven abilities, verbal ability has the highest correlation of the entire table, +.76, with the *Otis Test* The next highest correlation, +.68, also involves verbal ability with the *Pressey* It would appear from the correlations that both the *Pressey* and the *Otis* contain materials similar to those found in Thurstone's verbal ability It also seems that the materials of the reasoning, induction, and the perceptual factors overlap the materials of the *Otis* and *Pressey Tests*, as can be seen from the correlations. It may be that the perception ability is related to the speed factor found to operate in timed tests such as the *Otis* and

Pressey Memory ability correlates only slightly with the *Otis* and the *Pressey*. Induction and reasoning abilities correlate highest with the *Revised Minnesota Paper Form Board*. The primary abilities appear to correlate only slightly with the *Minnesota Assembly Test*. The *Minnesota Spatial Relations Test* correlates best of all with perception ability, but space and induction also appear to overlap to some degree with this test. On the *Minnesota Clerical Test*, *Name Checking* correlates highest with the perception and number abilities. On *Number Checking*, the five abilities of verbal, space, memory, induction, and deduction do not appear to be related On the other hand, four of these five abilities, with the exception of memory, appear to be related.

Finally, Tredick combined the abilities of number, verbal, induction, and deduction, using as her criterion semester point average, and obtained an R of +.61. Similarly, she combined the *Otis*, *Pressey*, *Minnesota Paper Form Board*, and *Number Checking* of the vocational guidance battery, using as her criterion semester point average, and obtained an R of +.57.

TABLE 7

Correlations for Each of the Abilities with the First Year's Semester Point Average in Engineering

Ability	P	N	V	S	M	I	D
Coefficient of Correlation	+.08	+.26	+.34	+.18	+.11	+.30	+.36

Goodman (7) made a second attempt in 1940 to determine the predictive value of the *Abilities Test* in determining college success for his engineer subjects at the completion of their first year in college Each of the factors was correlated with the first year's semester point average. The results are shown in Table 7

The abilities of reasoning, induction, verbal, space, and number were then used in a multiple correlation with first year's semester point average and the combined variables yielded a multiple correlation of +.49, which is slightly lower than that of +.51 obtained for the first semester

Goodman (7) then sought to determine whether he could

obtain better predictive values for his engineer subjects by using the sixteen Thurstone tests individually. These tests, when combined, yield the score that is the measure for the ability. According to Thurstone (14), each of the tests designed to measure the particular ability is highly saturated with that ability. The correlations between the tests that measure these abilities are shown in Table 8. The tests of the three abilities of number, verbal, and space appear to correlate highly with each other, while the correlations of the tests of the other abilities descend in size.

TABLE 8

Tests Measuring Each of the Abilities and the Correlations for the Tests of Each Ability

Factor	No.	Test	Coefficient of correlation		P.E.
P	1	Identical Forms	+.32		.05
	2	Verbal Enumeration			
N	3	Addition	+.72		.02
	4	Multiplication			
V	5	Completion	+.61		.03
	6	Same-Opposites			
S	7	Cards	+.63		.03
	8	Figures			
M	9	Word Number	+.35		.05
	10	Initials			
I	11	Letter Grouping	+.28	(11 & 12)	.05
	12	Marks	+.33	(11 & 13)	.05
	13	Number Patterns	+.22	(12 & 13)	.05
D	14	Arithmetic	+.44	(14 & 15)	.04
	15	Number Series	+.15	(14 & 16)	.06
	16	Mechanical Movements	+.01	(15 & 16)	.05

The abilities tests were then correlated with the first year's semester point average and the zero-order coefficients were obtained. For the purpose of comparison, the zero-order coefficients for the tests and the ability zero-order coefficients are given in Table 9 The following facts are to be noted: both tests of the perceptual factor correlate only slightly with the criterion as does the factor itself, but it can be seen that the *Identical Forms Test* alone correlates higher than the composite of the ability tests, while the correlation for the *Verbal Enumeration Test* is low in value. In number ability the r for the

single test of *Addition* is as high as the r for the factor itself and the *Multiplication Test* r is not much smaller In the verbal ability, the *Completion Test* r is again almost as high as the ability r. In space, the correlation for the test of Cards alone is greater than that for the ability, while the r for the second test of *Figures* is of low value. Both tests of memory correlate very slightly with the criterion while the r for the ability itself is slightly greater than the r for the *Word Number Test*. The three tests of induction have r's of approximately the same size, while the r for the ability is somewhat larger than that for any of the tests In reasoning, the r for the test of *Arithmetic* is almost as large as the r for the ability, with the *Number Series Test* r slightly smaller. The test of *Mechanical Movements* correlates slightly with the criterion and with the other two tests of the ability. Upon obtaining the zero-order coefficients between the tests and the criterion, a multiple correlation of eleven of the tests was computed The eleven tests were selected which correlated highest with the criterion and lowest with the other tests. The tests selected were *Arithmetic*, *Number Series*, *Same-Opposites*, *Completion*, *Letter Grouping*, *Addition*, *Marks*, *Number Patterns*, *Cards*, *Multiplication*, and *Word Number* The multiple correlation obtained, using these eleven tests, was +.48

TABLE 9

Zero-Order r's of the Sixteen Thurstone Tests with the Criterion

Factor	Test	Factor r with criterion	Test r with criterion
P	Identical Forms	+.08	+.09
	Verbal Enumeration		+.03
N	Addition	+.26	+.26
	Multiplication		+.24
V	Completion	+.34	+.31
	Same-Opposites		+.30
S	Cards	+.18	+.23
	Figures		+.10
M	Word Number	+.11	+.10
	Initials		+.08
I	Letter Grouping	+.31	+.26
	Marks		+.20
	Number Patterns		+.24
D	Arithmetic	+.36	+.32
	Number Series		+.32
	Mechanical Movements		+.10

It was apparent that there was little hope of raising the multiple coefficient by adding the remaining five tests, since they correlated very little with the criterion It will be noted that with the eleven tests the correlation was no greater than that obtained by using only five of the abilities

White (17) in 1942 made a study of the Thurstone tests and college prediction in Home Economics, using 94[a] of the same 113 subjects used by Tredick. At the time Tredick conducted her study these subjects were freshmen and had completed their first semester's work. When White conducted her study, the 94 subjects she used of Tredick's group were then sophomores. White combined the grades for all of the individual courses taken by her subjects and took as the representative score the students' average for each of the particular types of courses. Semester point average in White's study represents the average of two years of college work by her subjects Correlations were then obtained between each of the primary abilities and the average score for work in Art, Science, Home Economics, and semester point average. The results are shown in Table 10

It is worth comparing the results obtained by White as shown in Table 10 with those of Tredick in Table 4. White's data showed that each of the abilities correlated lower in all cases but one with semester point average than those obtained by Tredick. The exception is reasoning ability, which correlates +.45 with semester point average and indicates an increase over the correlation of +.42 obtained by Tredick.

The correlations in White's study for the abilities with Art average are generally higher than those found by Tredick. For Science average the r's of the abilities calculated by White are mixed, some being higher and others lower than those Tredick obtained for the abilities with Chemistry White's correlations are all higher with English average than those calculated by Tredick, with the exception of induction, which is slightly higher for Tredick All of White's *r*'s with Home Economics average are lower, with the one exception of induction, when compared with Tredick's correlations for Home Economics 101 and 109 Another fact worth noting is that in White's study all of the correlations of reasoning ability are higher than those obtained by Tredick. It may be that reasoning ability becomes more important when, presumably, the degree of difficulty of college work increases One more fact worth noting is that White's correlation of verbal ability with English average was + .65, while Tredick obtained an *r* of + .55 for verbal ability with English Composition

---

[a] Actually, the subjects are the same in each study. However, nineteen of the original group used by Tredick had dropped out of college, and since data were not available for them, White was unable to include them in her study

TABLE 10

Correlations of the Sixteen Primary Mental Abilities Tests and the Seven Abilities with Average Grades in Home Economics Courses

Ability	Tests	Semester average		Art		Science		English		Home Economics	
		1*	2†	1	2	1	2	1	2	1	2
P	Identical Forms	.05		12		.08		-.04		.02	
	Verbal Enumeration	.26	19	10	13	.23	18	.38	.20	.17	11
N	Addition	.33		14		44		30		19	
	Multiplication	.29	.33	.11	13	12	.44	.07	.28	14	17
V	Completion	48		.29		.32		.59		.33	
	Same-Opposite	42	49	.25	.29	.29	.33	.61	.65	.27	.32
S	Cards	12		.20		14		.07		.04	
	Figures	.27	19	.34	.28	.25	.20	.21	14	17	10
M	Initials	.20		14		.23		.28		.05	
	Word Number	15	.20	.04	.11	.26	.28	17	.26	-.10	.02
I	Letter Grouping	.28		.23		.26		.23		.21	
	Marks	12	.24	.14	.25	12	.23	14	18	.06	.14
	Number Patterns	.30		.23		.20		14		.09	
D	Arithmetic	43		.23		.51		.27		.35	
	Number Series	42	45	.29	.30	48	49	.31	.30	.27	.36
	Mechanical Movements	10		.10	.	.04		.06		15	

* Column 1 indicates *r* between tests and course.
† Column 2 indicates *r* between ability and course.

White then correlated each of the sixteen Thurstone tests with the average for each of the four courses and semester point average. These results are also shown in Table 10. Twenty-two per cent of the correlations for the single tests in Table 10 are higher than, or equal to, the correlations for the abilities In many cases the correlations of the tests of the ability are only slightly smaller than correlation for the ability Finally, White computed multiple correlations, using first, combinations of the abilities, and secondly, the individual tests When she combined number, verbal, memory, induction, and reasoning abilities with semester point average, she obtained a multiple correlation of + 59 Combining number, verbal, and reasoning abilities and correlating them with semester point average, she also obtained a multiple correlation of + .59. Lastly, she combined the five individual tests of *Completion*, *Same–Opposites*, *Arithmetic*, *Number Series*, and *Addition*, and obtained a multiple correlation of + .62

Other Thurstone Studies

A number of other studies have been reported on the predictive possibilities of the Thurstone *Primary Mental Abilities Tests*. Yum (19), in a study at the University of Chicago, calculated multiple correlations using various combinations of the abilities with semester average The best multiple correlation he obtained was + .422 using all of the abilities. Shanner and Kuder (10) have reported correlations of the abilities with average grades for the 1938 freshman class of the University of Chicago The highest correlation reported for average grades was with verbal ability, + .415. Correlations are also reported by these writers for the abilities with four introductory courses at the University of Chicago Deduction correlated with Biological Science, + .418; verbal ability, + 472 with Humanities, deduction, + 485 with Physical Sciences, deduction, + 427 with Social Sciences Multiple correlations using all seven of the abilities with the four introductory courses yielded the following *R*'s, + .500 with Biological Sciences, + .541 with Humanities, + .561 with Physical Sciences, + 566 with Social Sciences.

In 1941 Ellison and Edgerton (6) tested a group of 49 students at Ohio State University with the Thurstone *Primary Mental Abilities Tests*. The highest correlation obtained, + 44, was for verbal ability with point hour ratio The remaining ability correlations with point hour ratio range from – .24 to + .31. Combining the seven abilities, they obtained an *R* of + .640 with point hour ratio Ellison and Edgerton also report correlations for the abilities with grades in college subjects The highest correlations reported were, verbal ability with English, + 75, verbal ability with Science, + 68; induction with Foreign Language, + 78, reasoning with Psychology, + .63. They state "that the results can only be taken as suggestive and not as facts from which broad generalizations may be drawn"

On the basis of the studies reported in this paper, the following conclusions appear to be justified.

1 The Thurstone *Primary Abilities Tests* correlate, on the whole, as well as most standardized intelligence tests with criteria of college success.

2. The Thurstone *Primary Abilities* correlate with individual college courses to some degree and can be used for prediction of success in these courses

3. Verbal ability correlates higher than any other of the abilities with semester point average and individual college courses

4 Multiple correlations obtained by using various combinations of the primary abilities yield some increase in correlation over those obtained by using single abilities, when correlated with semester point average.

5. Multiple correlations using various combinations of the single tests that measure the primary abilities yield correlations with semester point average that were in some cases higher than or equal to those obtained using the abilities

6. Verbal ability correlates highly with the *Otis* and *Pressey* tests, and appears to be overlapping some of the functions in these general intelligence tests.

7 The *Otis* and *Pressey* tests appear to contain some of the same functions as those measured by the seven Thurstone primary abilities.

8. The single tests measuring an ability in some instances correlated highly with each other.

9 A single test of an ability will in some instances correlate higher with the criterion than does the composite of several tests of the ability itself.

REFERENCES

1. Ball, F J *A Study of the Predictive Values of the Thurstone Primary Mental Abilities as Applied to Lower Division Freshmen*. The Pennsylvania State College, 1940. (Unpublished thesis)
2. Bernreuter, R. G. *The Personality Inventory* Stanford University Stanford University Press, 1931.
3 Bernreuter, R. G "Primary Ability Tests Applied to Engineering Freshmen." *Psychological Bulletin*, XXXVI (1939), 548
4. Bernreuter, R. G. and Goodman, C H. "A Study of the Thurstone Primary Mental Abilities Tests Applied to Freshmen Engineering Students" *Journal of Educational Psychology*, XXXII (1941), 55–60
5. Chein, I. "An Empirical Study of Verbal, Numerical and Spatial Factors in Mental Organization" *Psychological Record*, III (1929), 71–94
6. Ellison, M L. and Edgerton, H A "The Thurstone Primary Mental Abilities and College Work." *Educational and Psychological Measurement*, I (1941), 399–406.
7 Goodman, Charles H. *Ability Patterns of Engineers and Success in Engineering School*. The Pennsylvania State College, 1941. (Unpublished thesis.)
8. Hessemer, Marianne. *The Thurstone Primary Mental Abilities Tests in a Study of Academic Success in the School of Chemistry and Physics* The Pennsylvania State College, 1942 (Unpublished thesis)
9. Schneck, M. M R "The Measurement of Verbal and Numerical Abilities." *Archives of Psychology*, XVII (1929), No 107
10 Shanner, W M. and Kuder, G F. "A Comparative Study of Freshman Week Tests Given at the University of Chicago." *Educational and Psychological Measurement*, I (1941), 85–92.
11. Stalnaker, J M. "Primary Mental Abilities" *School and Society*, L (1939), 568–572.
12. Strong, E K. *The Vocational Interest Blank* Stanford University Stanford University Press, 1938.
13. Tredick, Virginia D *The Thurstone Primary Mental Abilities Tests and a Battery of Vocational Guidance Tests as Predictors of Academic Success* The Pennsylvania State College, 1939. (Unpublished thesis)
14 Thurstone, L. L. *Manual of Instructions*. Washington, D. C.: American Council on Education, 1938.

15 Thurstone, L. L. *Primary Mental Abilities.* Psychometric Monographs, No 1, Chicago. University of Chicago Press, 1938

16. Thurstone, L. L. "Current Issues in Factor Analysis" *Psychological Bulletin*, XXXVII (1940), 189–236.

17 White, Elizabeth W *The Use of Certain Tests in the Prediction of College Success As Applied to the School of Home Economics.* The Pennsylvania State College, 1942. (Unpublished thesis)

18 Wolfle, Dael *Factor Analysis to 1940*, Psychometric Monographs, No. 3, Chicago University of Chicago Press, 1940

19. Yum, K. S. "Primary Mental Abilities and Scholastic Achievement in the Divisional Studies at the University of Chicago." *Journal of Applied Psychology*, XXV (1941), 712–720

TEST CONSTRUCTION IN PUBLIC PERSONNEL ADMINISTRATION

DOROTHY C. ADKINS

Social Security Board

Introduction

When the public is persuaded that positions in the public service should be filled by the best qualified persons and expresses its conviction through a civil service law, a tremendous responsibility, that of predicting which persons actually are the best qualified, devolves upon the agency charged with administering the law Increased emphasis on the impartiality of the selection of public officials has been accompanied by growing reliance on examining processes that are objective. Hence, the major attention of this article is devoted to topics relating to the construction of the written examination in civil service In this interpretation, problems common to the academic setting have been largely excluded. Although the article is further restricted to problems arising in state civil service or merit system jurisdictions, certain of the comments may apply equally to civil service at the federal level

However obscured it may be in practice, the essential of any merit examination is that it predict efficiency on the job Those who are not likely to perform satisfactorily on the job should be excluded from the final list of eligibles, and those who achieve places on the register should be ranked in the order of predicted job performance. To these ends each part of the total examination process should contribute

State civil service examinations often include, in addition to a written test, a rating of training and experience and an oral interview If proficiency in the operation of machines or equipment is essential, a performance test may be one of the components. For some positions phases of the examination may well be modified or omitted, but usually not the written examination

Rating of Education and Experience

Most state jurisdictions determine initially who shall be admitted to examinations by prescribing minimum qualifications of education and experience. Within limits, additional education may be substituted for a part of the experience, and vice versa. The process thus screens first those candidates who do not meet the minimum qualifications. Those barely meeting the education and experience requirements may be assigned the lowest passing score and those surpassing the minimum requirements higher ones, the scores depending upon such factors as amount, pertinency, and recency. The inclusion of ratings of training and experience as one part of the examination process is based on the assumption that differences among the candidates on these factors will be reflected in job performance.

The argument that the entire burden of the examination process should be placed on such ratings is untenable, however, since the rating procedure is relatively subjective and unreliable and fails to take cognizance of variations in the degree and extent of knowledges and abilities that exist even among those of similar training and experience. Such variations often are of greater significance in the prediction of job performance than are differences among candidates in education and experience beyond the entrance requirements Further, if these requirements are relatively high, differentiation among candidates meeting them may, by means of disproportionate weighting of mere survival, give undue advantage to those who have failed to progress.

Thus if, in relation to the salary level and labor market, minimum requirements for a professional class are low, say college graduation, then assignment of a high education-experience score to a candidate with two years of pertinent graduate work and three years of closely related experience probably would contribute to the validity of the total examination But if for a similar class entrance requirements are higher, say two years of pertinent graduate education and three years of experience, then giving a higher score to a candidate with five years of experience than to a candidate barely meeting the minimum qualifications may be of questionable value. The principal effect of such practice in this instance might be to give unwarranted higher scores to candidates who, although older, have not advanced beyond younger candidates also meeting the high requirements for the class Moreover, the rating of training and experience does not provide an evaluation of personality differences

The Oral Interview

The oral interview, despite its recognized limitations, seems to be the best available instrument for appraising personality characteristics For civil service examinations little or no reliance can be placed on paper-and-pencil tests of personality, which would fail to elicit frank answers in a competitive situation with jobs at stake Behavior in an oral interview can also be faked. For this reason, as well as unreliable rating, the oral interview included in the final score is usually weighted considerably less than are the other two parts, even though it contributes positively to validity, and the number of candidates failed on the oral is very small.

Warranting at least passing mention are certain misconceptions of what constitute desirable purposes of an oral interview, such as the idea that its aim is to reappraise the applicant's education and experience or that it should be designed to ascertain the scope of the candidate's knowledge or the degree of his general intellectual abilities. Since such factors can be measured more adequately by the other parts of the examination process, the inclusion of a relatively unreliable rating of them in the total score probably serves to lower not only the reliability of the total composite but, more critically, its validity.

The Performance Test

Where differences in skill in operating machines is a pertinent factor, the total examination usually includes a performance test. For positions requiring such skill, the oral interview is of negligible importance if the duties typically entail little contact with the public or with fellow employees Because of this and not because they serve similar purposes, performance tests and oral interviews are rarely used for the same class Differential ratings of training and experience are also commonly omitted for machine-operator classes Performance tests are costly, difficult to administer, time-consuming, and frequently not very reliable. But their use may add appreciably to the predictive efficiency of the composite. In view of their limited reliability and their failure to take adequately into account the ability to profit from further training and experience, they are more appropriately regarded as a supplement to rather than a substitute for a written test.

The Written Test

Test Content. Even though the other parts of the total examination process may be so important for certain classes that they should not be dispensed with, no other is so significant or should bear so much weight as a well constructed written examination. The purpose of the written test is to determine reliably the extent of individual differences in pertinent areas of knowledge and abilities Defining areas of knowledge and the abilities to be sampled in an examination is achieved by means of job analysis, the results of which are commonly summarized in a class specification Such a compendium, depending upon its thoroughness and clarity, gives a picture of the duties of the class of positions, qualifications in terms of education and experience, the supervision exercised and received, and the knowledges and abilities bearing on the duties of the class. This information should be supplemented by knowledge of the relationship of the class to the total organization and of the applicable salary range First-hand acquaintance with the job, although unfortunately not always available, is of inestimable value. Since the written examination must contribute to the prediction of efficiency on a particular type of job, its most effective construction is contingent upon a clear idea of what the job is. Its subject matter should first of

all be related to the prediction of job performance; and the weights to be assigned different areas of subject matter should depend upon their contribution to the prediction of efficiency on the job

The types of subject matter included in a test have all too often been limited by lack of facilities essential for item construction Administrators and test technicians place too little reliance on persons thoroughly familiar with the subject matter. Test technicians sometimes delude even themselves into believing that they can construct or at least assemble an adequate examination in a specialized field without assistance. Much of the criticism directed at written tests is attributable to action based upon just such misguided self-confidence. On the other hand, examining agencies should not vest sole responsibility for either item construction or assembling examinations in subject-matter consultants who are not skilled in examination techniques The sounder and more successful approach assumes collaboration between those schooled in content and those versed in technique, although an occasional agency may be fortunate enough to have on its staff a person thoroughly competent in both a subject-matter field and examining procedures

The content of the written examination should be related to realistic class specifications It is by necessity limited by the facilities for constructing or securing items A third factor that cannot be ignored in its bearing on examination content is public opinion. Almost universally, civil service examinations are required to be "practical and related to the job," although the exact statement of the criterion varies It would seem at first thought that establishment of a "satisfactory" relationship between test scores and performance on the job would guarantee the meeting of this criterion—and the merits of such an argument will not be denied. Pertinent to this contention, however, is the difficulty of establishing that the correlation of an examination with job performance is "satisfactory," particularly in such a way that the public will accept the demonstration as proof of the "practicality" of the examination. Faced with this dilemma, the majority of civil service agencies

interpret the term practical to mean what they think the public means by practical. A "practical" examination, then, becomes one that *looks* practical to the lay person, and, particularly, to the candidate An examination may of course be satisfactorily valid from the point of view of its correlation with job performance and yet fail to meet the criteria for "face validity" of this type. In the interest of fostering public support of the civil service principle, construction of examinations efficient in predicting job performance and also acceptable to the public as practical is the desideratum. An examining body that limits itself to the kind of face validity under discussion is not in any sense adequately fulfilling its function; nevertheless, its general aim will in the long run be appreciably advanced if it achieves discriminating examinations that, at the same time, have this type of validity

Rigorous application of this criterion of practicality requires not only careful scrutiny of the broad areas of knowledge and abilities sampled in the test as a whole, and of the general area covered in each item, but also attention to the several individual concepts contained in each item Sometimes attention must be focused on the individual word In a test composed of multiple-choice items, for example, the candidate may not recall merely the question or just the question and the best answer—obviously, he may not know which answer is intended to be best But he may remember and criticize some of the answers intended to be "distracters," "decoys," or "confusions," as they are variously termed. If an item constructor innocently includes Socrates in a distracter, he should not be surprised if the examination is later publicized as absurdly impractical because it inquires into the candidates' Greek philosophy Particularly in the case of civil service agencies not far removed from the publics they serve, every examination item should be reviewed from the standpoint of how it might look in the public prints

Speed versus Power Public opinion also has important bearing on other aspects of the examining process. For example, it is more favorable to a power test than to a test that places a premium on speed.[1] People say job duties are not performed in a setting that emphasizes speed and competition.

These in turn, they say, create anxiety and thus poor performance. The public subscribes to the belief that "accuracy is more important than speed" without recognizing the corollaries that for many jobs, rapid is preferable to slow accuracy, and that speed and accuracy, at least in relatively simple tasks, tend to be positively correlated

Many candidates, and particularly those who have failed or attained low ratings on a time-limit test, share the layman's preference for work-limit tests Some may believe that speed tests unduly penalize the older, more experienced worker The fallacy in this argument is twofold first, although speed of performance of *some* functions, as measured by tests, probably does decrease somewhat with age, in reality the speed of the performance of the *same* candidate at *different* ages (within the usual age range of candidates but excluding the age of senility) does not usually differ markedly; and, second, if the speed of performance of particular functions *did* decrease appreciably with age, then, for those positions for which speed of performance of these functions is important, reflection of this decrease in the test score would increase the validity of prediction. Actually, comparisons referred to by candidates are usually not based upon the performance of the same candidates at different age levels but rather upon that of persons who differ not only in age but also in basic abilities and who still would differ in abilities even if they were of the same age The older candidates may get lower scores on a speed test and thus appear to be penalized The score is not a result of age but an indication of lesser ability. Reflect that most older candidates for a job at a level for which a speed test is frequently given have worked for a number of years without progressing as far as jobs which will be beginning jobs for the younger candidates Thus, among candidates for clerical positions, those aged 55 do not represent the same kind of sample of 55-year-olds as those aged 19 represent of 19-year-olds.

Regardless of the false premises underlying this predilection for power tests, probably the best forecasts of performance in the higher-level positions will result from tests with time limits so liberal that only habitual laggards will have difficulty in attempting all the items within the time allotted. For the majority of clerical positions, however, where speed of performance clearly is an important component of job proficiency, achievement of the best possible prediction now will in the long run outweigh any advantage that might be gained by a concession to public opinion at the moment The inclusion of a speed test as one component of an examination, even for a clerical class, may wisely be accompanied by a campaign to educate the interested public on the reasons for its use.

Length. Every person who has ever attended school has opinions about examinations. Thus public opinion impinges on the proper length of a test. The school examination, taking from 30 to 50 minutes, is regarded as typifying the most trying ordeal to which a candidate should be subjected It is forgotten that normally the results of any one such school examination are considered along with the results of several others, daily observation, and the appraisal of performance on many additional assignments, and, more important, that such an examination is designed to measure achievement in a relatively limited area of a single subject-matter field. In contrast, the civil service written test is supplemented, if at all, only by a rating of education and experience and possibly by a brief oral interview or a performance test; and it is designed to sample a much greater complexity of knowledges and abilities

When candidates clearly realize this distinction, when they understand why a test that adequately samples the major fields relating to a job is more reliable and more valid than one restricted in coverage, they prefer tests sufficiently long to yield a reasonably just appraisal of their job potentialities Although candidates may complain that long examinations are endurance contests, experimental work on prolonged mental effort indicates that the *affect* is more potent than the *effect*. Unfortunately, some civil service jurisdictions capitulate to public pressures for short examinations instead of enlightening the public on the virtues of comprehensive sampling

Type of Item Decision on the type of item best adapted to civil service tests should be weighed against adequacy of

coverage of the pertinent areas of knowledge and abilities, objectivity of scoring, and ease of administration. All things considered, the use of a large number of objective items is in general preferable to reliance on an essay examination. In civil service, objective tests have largely superseded essays In agencies that still use them, the tendency towards supplementation by the objective type increases

If broad, general essay questions are used in an effort to cover particular fields of subject matter, different candidates may not actually be answering the same question, a factor that may nullify attempts to place them in an order of merit If, on the other hand, essay questions are more pointed and limited in scope, the coverage of the requirements is automatically restricted, although more reliable grading can be achieved The objective test not only has the advantage of permitting a broader sampling of pertinent knowledges and abilities than the essay test but also, if properly used, will almost certainly lead to markedly greater reliability in scoring. These are highly important advantages in civil service, where areas of subject matter to be covered are broad and where unreliability of scoring may have a greater effect upon human destinies than in almost any other field of testing.

Test scoring should be objective and also as simple as possible in the interests of efficiency. Wherever feasible, the same kind of item should be used throughout and the same weight assigned to each item. The multiple-choice form lends itself admirably to most purposes. Many forms that appear to differ are simply variants of this form. The argument that use of a variety of forms adds interest to a test is insignificant, since the competitive setting supplies more than enough motivation Arousing interest in a test per se is an empty gesture. Items, and even the several responses to an item, are sometimes differentially weighted in civil service tests, but there is growing awareness that unweighted and weighted scores for a large number of items correlate so highly that little is to be gained by differential weighting.

Use of items of a single, readily understood type, accompanied by clear instructions to candidates and with one over-

all time limit, simplifies test administration and hence may contribute positively to test validity. This is one way of overcoming the handicap of monitors poorly trained or too unstable to meet emergencies with poise and skill.

Repeated Use of Items In the educational world test questions are sometimes used more than once, to the gratification of certain fraternities that keep files of questions available Scores tend to be higher on the second administration of a test. Perhaps in college repeated use is not very serious, since a course mark usually does not depend on one examination alone and since a mark in a single course may be of no great moment. Institutions that are placing dependence on comprehensive examinations and little if any on course marks, however, are increasingly following the practice of using test questions only once. The problem of maintaining the confidential nature of examination materials is even more urgent for civil service jurisdictions There may be organized efforts on the part of "cram schools" to obtain access to examination items used, some candidates may apply for the sole purpose of memorizing assigned portions of an examination In the absence of such studied attempts to sabotage a merit system, candidates nevertheless may remember a few items in detail and another group of questions to "look up." Whether or not this factor would give any odds to these candidates or their friends if the same items were repeated, some members of the public might *think* there would be an advantage. Thus the confidential nature of the examination fuses with the question of public relations. A civil service agency better maintains the support of the public if it can give assurance that no appreciable advantage *could* accrue to any candidate because of previous use of items.

Statistical Analyses of Items. From the point of view of the reliability and validity of a test, inclusion of a few poor items is negligible if the test as a whole is long enough. In one way, however, the problem created by indefensible items is greater in civil service tests than in almost any other kind This is true because a difference of one point may affect a candidate's rank order and also may determine his passing or failing. Hence he may or may not get a job. Since a difference

of even one point has this importance, great care should be exercised to exclude from a test items that might have a negative validity The problem is of consequence, too, as it relates to public faith in the merit system A widely heralded appeal based on a few weak items, which may really be of no great moment insofar as they affect the over-all reliability and validity of measurement, can go a long way toward undermining public support of the merit principle.

In speaking of the validity of a civil service test, we should have in mind the extent to which the test serves its basic purpose, prediction of performance on the job To establish incontrovertibly that a test has validity for this purpose is of course difficult. If a candidate population is used, the failing candidates do not get jobs, so that there is no information regarding their work, and the distribution of job performance indices is thus curtailed to an indeterminate extent. Even more troublesome is the notorious unreliability of service ratings commonly used to appraise success on the job Made by a number of raters, with individual standards, on employees performing for varying periods on jobs that differ although grouped under one classification, service ratings are sensitive to many uncontrolled and uncontrollable factors Too frequently clearly recognizable differences in job performance are obscured by a tendency to rate employees in the "above average" categories, with the result that distributions of ratings exhibit marked negative skewness And this is only one of the frailties to which the rater is heir! Unfortunately it is almost impossible to state just how unreliable service ratings are, since conditions for a really crucial experiment in this area have not been met in practice. Probably, however, .30 to .50 represents a safe estimate of the Pearson correlation coefficient between two completely independent sets of ratings, made by raters assumed equally familiar with the work of employees on jobs within the same class, using rating forms developed with an ordinary amount of skill and care A similar correlation of ratings made specifically as a criterion against which to judge a test would tend to be higher. In any case, problems of securing as a criterion reliable evaluations of job performance on a population of sufficient size are very great and in most instances insurmountable.

Establishment of test validity for a particular group of candidates is troublesome enough, but efforts to prevalidate on one group a test to be used on another group for selective purposes are more hazardous still

If a population of employees rather than candidates is used, the problem of proportionate representation of those who would fail the test if they were candidates remains Assurance is lacking that an employee group, for purposes of a validity experiment, is representative even of the passers among a candidate group Parts of a test that are satisfactorily valid for a candidate group may not differentiate among employees, who have had experience on the job and who may all have learned to do certain kinds of tasks that only the abler of the candidates could perform Moreover, if the confidential nature of tests is to be maintained, an agency may not wish to be in the position of repeating, either wholly or in part, a test from which there may have been "leakage."

Attempts to secure an experimental group of subjects who are neither candidates nor employees face these difficulties as well as the impossibility of obtaining any measure of job performance.

Confronted with these obstacles, the experimenter in the field of civil service tests usually resorts to a candidate population and an internal criterion, which is customarily the total score on a test composed of items sampling a variety of areas of knowledge and abilities Conclusions should be drawn only charily from analyses based on such a multifactored criterion. Since different areas are represented in the variance of the total scores to differing and usually unknown extents, dependent upon the variances and interrelations of the part scores, generalizations on relative validities are apt to be inappropriate and misleading. The item-test coefficient must be interpreted in relation to the item difficulty for the population in question. An item that fails to differentiate among candidates for one type of position may discriminate positively for another Any items with negative coefficients should be carefully scrutinized with a view to improving the item if its re-use is contemplated and, in any event, to gaining insight into the characteristics that may

lead to negative coefficients An item that correlates negatively with an internal criterion may, however, be positively related to the ultimate criterion to be predicted.

For multiple-choice items, it is advantageous to examine not just the item-test coefficient, which may be regarded as a kind of summary statistic, but also the relationship of each option to the criterion This has been approached in various ways—by correlating each choice with the criterion, by finding the mean criterion score of those who select each choice, by finding the proportion of persons in each of two contrasted criterion groups for each choice, or by some other method, depending upon the type of item index preferred With access to this kind of information for each choice, the examiner can often improve items and learn how to construct better ones Within certain limits, which of the many variant forms of item-analysis techniques one uses seems relatively unimportant. Ideally, it may be preferable to compute for each choice for each item a correlation coefficient (either a biserial, a point biserial, or both) with a multivariate criterion[1] If interest centers in only a small number of items, this approach is entirely practicable. If, however, results on several hundreds or even thousands of items are available, a degree of statistical refinement may profitably be sacrificed in the interests of using a larger proportion of data at hand, particularly if facilities for such research are limited. In this case, a reasonably satisfactory technique may be to use the tetrachoric correlation coefficient with the criterion dichotomized at the median. Here, again, it will be of value to find in addition the proportions of candidates above and below the criterion median who select each choice For most purposes, a practical *index of item difficulty* is simply the percentage of candidates who select the "best" answer

Some agencies that do not at the present time want to repeat items nevertheless apply item analysis techniques with a

---

[1] Since normally it is not intended to repeat any large group of items from a given test, there is no advantage, and some disadvantage, in the application of any of the "build-up" methods, such as those of Toops, Horst, and Richardson, in analyzing civil service tests. In any case, these methods are designed for use with an external criterion, although the first two may be modified for an internal criterion.

view to gaining insight into the characteristics of good and poor items. Both subject-matter consultants and test technicians will profit from reviewing results of statistical analyses, even though the values may be in large part indirect

As data accumulate over a period of years, it should be feasible, at least for classes attracting large numbers of candidates, to construct an examination largely from pretested items only a small number of which have been used before in any one examination. The results of statistical item analysis should not be applied blindly, for both the reliability and validity of items may be altered strikingly by changes in the social milieu. A test item valid at one time may be based on a concept that later becomes such common knowledge that even the poorest candidate answers it correctly Or it may hinge on a concept that becomes outmoded For such reasons as these, even were statistical indices available for a group of items from so large a number of different tests that the confidential nature of examinations would not be jeopardized by judicious repetition, competent consultants' review of items shortly before their inclusion in a test would still be essential to sound test construction.

Perhaps in the not too distant future factor analyses of civil service tests can dispel the problem created by the unknown contributions of several factors to the variance of internal criterion scores Then items can be correlated with separate factors instead of with a hodgepodge. Even so, the question of the appropriate weight for each factor in the composite could not be answered rigorously. The solution to this problem awaits development of a satisfactory external criterion

Establishment of Critical Scores As those experienced in test construction know, even when statistical data are available the difficulty of an item cannot usually be predicted with high accuracy unless one has worked a great deal with the particular type of item and is predicting the difficulty for a group of candidates having a known level of ability on the type of item in question Although the prediction for each item may not be accurate, many civil service agencies assemble groups of items with a view to setting the passing point on an *a priori* basis.

In some cases, the law or rules under which the agency operates make such a judgment mandatory In many others, where a fixed passing point is not so specified, the agency nevertheless considers the predetermination of passing points desirable, primarily from the point of view of public acceptance of the 70% passing point via the educational system Truce with this concept, hoary with tradition though it is, is not entirely without compensation.

Although the judgment of the appropriate difficulty of a test is admittedly hard to make, adherence to a fixed percentage of the total number of items as the passing point, where its use is not clearly inappropriate to needs of operating agencies or unjust to candidates, puts the examining agency in a position to combat pressures for setting passing points so that particular candidates are passed. Thus the agency is better able to be impartial Not infrequently, however, even in agencies that have a fixed percentage passing point set by law, scores are transmuted if an examination appears to have been so difficult that a register resulting from it would contain insufficient names to fill vacancies Usually a linear transformation is applied so that some percentage score less than 70 becomes a derived score of 70 and, obviously, so that some other condition is satisfied, such as the original top score being equated to a score of, say, 95 on the derived scale The most appropriate second condition varies with the character of the original distribution and with the desired properties of the distribution of derived scores

Although it may be considered advantageous to attempt to estimate test difficulty so accurately that no transformation of scores will be necessary, transmuting upward is considered preferable to transmuting downward, because candidates, while welcoming scores seemingly higher than actually attained, are reluctant to accept scores lower than the percentages of items answered correctly Moreover, a test that is difficult will discriminate among candidates better than one that is too easy

Several factors in addition to the inherent complexities of estimating test difficulty magnify the problem for the civil service examiner. One is that the nature of the candidate population varies with the labor market, and the extent of change, as it pertains to the abilities and knowledges being measured, is not easy to forecast. It varies with recruitment, again unpredictably. It varies with changes in minimum qualifications and salary level, in such a way as to be especially bothersome if minimum qualifications are lowered at the same time that the salary level is increased, a combination not unusual in a tight labor market. In fact, with the multiplicity of factors operative, it is surprising that results as satisfactory as those commonly achieved are possible.

Sometimes effort is made to set the critical score at a "break" in the distribution, on the grounds that failing candidates will accept their lot more readily. Some evidence to bolster this point of view probably could be adduced. At times, however, the procedure seems to be based on the assumption that peculiar significance attaches to a segment of the range within which no scores fall. This is nonsense. It is doubtless possible to construct a test having scores with a bimodal distribution. Whatever gaps appear in distributions of civil service test scores, however, are due to chance, not to rational design. If the number of candidates is sizeable, there will be no appreciable breaks in the proximity of any reasonable critical score, and no matter where this score is set there will be scores just below it. All things considered, an agency should face this fact tough-mindedly and not waste time in seeking breaks in score distributions or in creating them artificially.

Compiling Related Examinations. A task almost peculiar to civil service examination construction is that of assembling for a single administration examinations for sometimes as many as 35 or 40 classes, but perhaps more typically for from five to 15. If there are no common requirements of knowledges or abilities among classes for which examinations are to be held, then the examinations are simply assembled independently. If there are some requirements common to two or more classes the examinations ordinarily include some common items sampling the overlapping ones, particularly identical degrees of a given kind of knowledge or ability.

One argument for using overlapping items for related classes

is that some candidates may want to take more than one examination; if the extent of overlapping is slight, such candidates may not complete some of the examinations in the time allowed. To permit candidates to take several examinations for related classes in a single day (usually the maximum time for administration of an examination program), the only alternative to overlapping is shortening the total length of each examination, which is of course disadvantageous from the point of view of reliability. Some jurisdictions err in the other direction by constructing tests with so large a proportion of overlap that differentiation among the tests is unreliable. The effect of this error is that candidates who fail one examination are likely to pass another for a higher class in the same series, because of the chance element in the small number of differentiating items.

Such a result is not the *reductio ad absurdum* of the examination process that it may seem at first glance. As a matter of fact, identical examinations for classes that differ in degree rather than in kind would in some instances be appropriate were it not for interpreting to the public several passing point standards. On the other hand, the same passing points might be set for the written component of the examinations for several classes in a hierarchy, if the total examinations for the classes were differentiated on some other basis. Such a practice has sometimes been followed by the United States Civil Service Commission. For jurisdictions more limited in size of the public served, different written examinations for separate but related classes seem preferable, because some of the candidates for each are likely to be placed in the same agency on the basis of the examinations.

Roughly stated, the most useful principle for differentiating examinations for classes in a series is that the examinations should contain enough different items to reflect reliably the dissimilarities in requirements and enough common items to enable candidates who so desire to take three or four examinations at one sitting. A further important advantage in the use of common items is the economy effected in item construction and review. Although there is no rule applicable to all cases, prob-

ably the *minimum* number of differentiating items for two examinations for closely related classes, each consisting of 200 items, should be in the neighborhood of 50 to 70 items. The *optimum* number differs considerably from one pair of classes to another.

Order of Items for Related Examinations. A special problem of ordering items arises when, say, 15 examinations are encompassed in a booklet of, say, 1000 items, many of which are common to two or more classes. If they are ordered strictly according to subject matter, the items in any one examination are scattered, hence candidates who have to skip several groups complain that the mechanics are confusing and may waste their time. On the other hand, if the items are ordered in such a way as to minimize the number of "breaks" for the examinations having the largest numbers of candidates, the items can no longer be ordered entirely by subject matter, nor can an ordering from easier to more difficult be strictly maintained.

If administrative controls are adequate, perhaps the best solution is to assemble for each candidate a booklet containing only the items pertaining to the examination or group of examinations he is taking. Such a plan creates limitless confusion if the program is improperly administered.

A solution that places a lesser burden on sound administrative controls is to retain the plan of assembling related examinations in one booklet and to compromise among the objectives of arranging items (1) in the order that minimizes breaks in each examination, (2) in the order that seems logical so far as subject matter is concerned, and (3) in order of difficulty. Perhaps the simplest way to arrive at an ordering reasonably satisfactory from these three points of view is first to establish an order that minimizes breaks for the most "populous" classes, then to adjust the ordering by subject matter, then further to rearrange the items so that for any single class the more difficult items in any distinguishable area follow the easier items.

Selection and Training of Subject-Matter Consultants. Earlier emphasis was given to subject-matter consultation in civil service test construction. Brief mention will now be made

of some of the problems of using consultants in this capacity. Both selecting and training consultants are simplified for those examining agencies needing full-time consultants and having money to pay for them. More typically agencies seek intensive but intermittent services.

Unfortunately those who are recognized as authorities in their fields are also those who are likely to be employed in positions from which release for temporary assignment elsewhere is not easy to obtain. On the other hand, even if an examining agency has a continuing job of such magnitude as to warrant a permanent assignment of a consultant, many of those who would be acceptable as consultants are loathe to abandon front-line operations to undertake the task of predicting the performance of others. They hesitate to leave a position that they know and like to pioneer in a job that they may find difficult to understand and that may offer no clear line of advancement. Having taken this hurdle, they become dissatisfied if opportunities for contacts with others in their field are too limited or if for some other reason they "go stale" on the job. Because of these factors, some agencies tend to prefer making a series of temporary appointments, possibly on a part-time basis.

Probably no single solution to the problem of selecting and training subject-matter consultants would meet all needs. Where it is appropriate and feasible, however, a plan of having a senior consultant on a full-time and permanent basis and additional consultants on a part-time or temporary basis has the advantage of providing continuity to the examination program while at the same time securing a reflection of different points of view in the examination content. This plan also simplifies the training problem, because a permanent subject-matter consultant who understands examination construction can interpret aspects of this field to other consultants in the same area, often more efficiently than can the psychometrician.

Some persons who are sought as consultants and who may be recognized as authorities simply cannot be taught to construct examinations. Sometimes the difficulty is apparently attributable to temperament, sometimes to pattern of abilities, more often to the combined influence of both factors. To decide that a consultant cannot be taught to construct items may require only half an hour and again, for one who approaches the task with interest and enthusiasm, may take a month or more Such ineptitude is not peculiar to any field It can be noted in psychology, art, grammar, accounting, social work, law, music, and statistics The fruitful solution to this problem is to get another consultant as tactfully as possible.

Once a consultant is found who is both interested in and adept at item construction and test compilation, he should be given interpretation on the desirable length of an examination, scoring procedures, time limits, reliability and validity, transmutation of scores, weighting of several components, and similar concepts. Two of the most common misconceptions of subject-matter consultants are confusion between the difficulty and the validity of an item and failure to differentiate between a test as a predictive instrument and as a teaching device. Much attention must be given to overcoming these and similar fallacies related to examinations

Magnitude of the Task The final problem of which brief mention will be made is the scope of the examining job to be done within budgetary limitations. Although monies allocated to the examining function in any area are rarely of staggering proportions, probably the public personnel agencies suffer the most from paucity of funds The number and variety of jobs for which examinations are to be constructed and the sizes of the populations to be examined in normal times are in many instances almost unbelievable. In a single jurisdiction hundreds of lives may be affected by a single examination program Yet the annual budget for the examining function probably would represent only a small fraction of the sum allocated yearly to the "supervision of reindeer in Alaska." Despite the handicaps of inadequate staff and limited facilities for research, the notable progress of the last decade places on those in the field of measurement an obligation for continued effort toward the solution of the critical problems that remain

RELATIONSHIP BETWEEN KUHLMANN-ANDERSON INTELLIGENCE TESTS IN GRADE 1, AND ACADEMIC ACHIEVEMENT IN GRADES 3 AND 4 *

MILDRED M. ALLEN
New Rochelle Public Schools, New Rochelle, New York

This study is part of a doctoral research on the prediction of academic success of elementary school pupils by means of the *Kuhlmann-Anderson Intelligence Tests*. The purpose of this study was to determine the predictive value of the *Kuhlmann-Anderson Intelligence Tests* as a whole when administered in Grade 1, in the fields of reading, arithmetic, and spelling in Grades 3 and 4 as measured by the *New Stanford Achievement Test*.

Data

The subjects for this study were three hundred and twenty-seven pupils from ten elementary schools in New Rochelle, New York. Complete test results of these pupils from the school years 1936–37 to 1939–40 were used. The tests were the *Kuhlmann-Anderson Intelligence Test* and the *New Stanford Achievement Test*

Procedure

An alphabetical class list of fourth-grade pupils (1939–40) by schools was obtained from the school census clerk. From this list pupils were selected who were originally in the first grade in 1936–37. A checking and re-checking of all test results dating from Grade 1, 1936–37, and including the fourth grade of 1939–40 was made in order to select a group of pupils who had taken the complete battery of tests as used in the present study

---

* Part of a study for a Doctor's dissertation completed at New York University, Graduate School of Education, 1940.

The *Kuhlmann-Anderson Intelligence Test for Grade 1 (Second Semester)*, was administered in February, 1937, when the pupils were midway through the first grade. The *New Stanford Achievement Test (Primary Examination) Form Z*, was administered in April, 1939, near the close of the third grade, and the *Advanced Examination, Form W*, of the same test was administered to the same pupils in the fourth grade in October, 1939.

The *Kuhlmann-Anderson Intelligence Test* was personally administered, scored, and re-scored by the writer The *New Stanford Achievement Test (Primary Examination)* was administered in the elementary schools by the principal or a teacher who had test experience and training. Both principal and teacher-examiner received instructions for the administration, scoring, and tabulation of test results from the writer The tests were scored, and double-checked by specified teachers in the respective schools, by the principal, or by the teacher-examiner. The *New Stanford Achievement Test (Advanced Examination)* was administered by the writer to all fourth grades in October, 1939 This test was scored and re-scored by an assistant under the direct supervision of the writer.

Results and Interpretations

The mental ability of the pupils used in this study is shown in Table 1 Correlation coefficients between measures obtained from the *Kuhlmann-Anderson Intelligence Test* in Grade 1, and performance on the *New Stanford Achievement Tests* in Grades 3 and 4 are shown in Tables 2 and 3

TABLE 1

Means, Sigmas, and Ranges of I.Q.'s on the Kuhlmann-Anderson Intelligence Test *in Grade 1 and in Grade 4*

Test	Mean I.Q.	σ	Range
Kuhlmann-Anderson Test-Grade 1, Feb., 1937	100.7	9.9	63–125
Kuhlmann-Anderson Test-Grade 4, Oct., 1939	99.8	11.9	67–132

The mean I.Q.'s of the pupils indicate average ability for the group as a whole, with the range of I.Q.'s showing the wide

scatter of ability commonly found in heterogeneous grouping. The intelligence level of the pupils remains about the same in Grade 4. (These were the same pupils tested in Grade 1)

TABLE 2

Coefficients of Correlation Between Kuhlmann-Anderson Intelligence Test *Performance in Grade 1 and the* New Stanford Achievement Test (Primary Examination, Form Z) *Performance in Grade 3*

New Stanford Achievement Test, Grade 3	Kuhlmann-Anderson Test—Grade 1		
	M.A.	I.Q.	Pc.Av.*
Paragraph Meaning	40	44	.43
Word Meaning	.32	.39	40
Reading Average	.37	42	.42
Arithmetic Reasoning	48	.45	.45
Arithmetic Computation	45	43	.43
Arithmetic Average . . .	.52	.50	.49
Spelling	.36	40	.42
Total Average	.46	.49	.49
Educational Age	.48	.53	.53
Educational Quotient . . .	.40	.67	.67

* The Pc.Av., or Per Cent of Average Development, is an index obtained by dividing an individual's mental unit points by the average mental unit points for his age group, mental units being determined by conversion to a point scale designed by Heinis The Pc.Av is preferred by Kuhlmann to the I.Q., since it is more constant for retests over a period of years.

Table 2 reveals coefficients of correlation between the *Kuhlmann-Anderson* measures and educational achievement ranging from .32 to .53, with twenty of the thirty coefficients between 40 and .50. In view of the fact that these coefficients have standard errors of from .03 to .05, no significant differences among the coefficients are indicated. The I.Q., Pc.Av., (Per Cent of Average Development) and E.Q. all include one common element, namely, the chronological age (C.A.), which is not included in the mental age (M.A.) score It may be noted that some persons regard as spurious correlations of ratios involving the same variable denominators[1] The critical ratio for the differences of the coefficients of .53 (I.Q. or Pc.Av., and E.A.) and .67 (I.Q. or Pc.Av., and E.Q.) is 3.1.

Whether one considers the M.A., I.Q., or Pc.Av seems to make little difference since all yield substantially the same coefficients with the various subtest results on the *Stanford Pri-*

---

[1] J. P Guilford, *Psychometric Methods*, p 374.

mary Achievement Test. One part of the *Stanford Primary Achievement Test* is predicted as well from the various *Kuhlmann-Anderson* measures as any other part. The only variable derived from the *Stanford Achievement Test* which shows any significant differences in the coefficients is the E.Q. An examination of the last row of Table 2 shows that the correlation between M.A. and E Q is about the same as the other coefficients in the table The relationships between both the E Q and I.Q., and E.Q. and Pc.Av, are substantially higher (40 compared with .67 and 67) The critical ratio of the difference between the correlation of M A and E.Q, and the correlation of I.Q. and E.Q. is 5 4 This indicates undeniable statistical significance of the difference.

Excluding the coefficients involving E.Q., the median coefficient for Table 2 is 44 While this indicates some degree of relationship between the *Kuhlmann-Anderson* scores and educational achievement, it is by no means great enough to be of much value in predicting third-grade performance from the *Kuhlmann-Anderson Intelligence Tests* administered in Grade 1. Due to differences in content material of the earlier *Kuhlmann-Anderson Test* (non-verbal) in Grade 1, and the verbal content of the *Stanford Primary Achievement Test* in Grade 3, it appears that long-range predictions of academic achievement would not be reliable To insure a greater degree of reliability, intelligence tests should be repeated annually, and preferably within the grade for which predictions are made. The coefficient of alienation corresponding to an r of .44 is .8980, an indication that errors of prediction would be reduced by only 10.2 per cent by the use of the *Kuhlmann-Anderson Intelligence Test* in Grade 1, instead of making predictions without the tests. The highest coefficient of Table 2, .53, between I Q. (or Pc Av) and E.A. yields a k (coefficient of alienation)[1] of .8542, indicating a reduction in errors of prediction of only 14 68 per cent better than chance

The most predictable measure obtained from the *New Stanford Achievement Test, Primary Examination,* when predictions are made from the *Kuhlmann-Anderson Intelligence Test* administered in Grade 1, seems to be the Educational Quotient (E.Q) The r of 67 between I.Q (or Pc Av.) and E Q. yields a k (coefficient of alienation) of 7424 and indicates a reduction in error of prediction of 25 76 per cent better than chance The E.Q , however, does not give any indication of the actual level of achievement, but only of achievement compared with C A. It is significant, however, as an index of educational brightness In this study the Pc.Av (Per cent of Average Development) does not appear to have any advantage (for predictive purposes) over the I Q or M.A scores obtained from the same test

---

[1] J P Guilford, *Psychometric Methods*, p 362

TABLE 3

Coefficients of Correlation Between Kuhlmann Anderson Intelligence Test *Performance in Grade 1 and* New Stanford Achievement Test (Advanced Examination, Form W) *Performance in Grade 4*

Stanford Achievement Test—Grade 4	Kuhlmann-Anderson Test—Grade 1		
	M.A.	I.Q.	Pc.Av
Paragraph Meaning	.37	41	.43
Word Meaning	.30	.36	.35
Reading Average	.35	42	41
Arithmetic Reasoning	.52	.51	.50
Arithmetic Computation	.49	46	.49
Arithmetic Average	.56	.53	.53
Spelling	.36	.39	.42
Total Average	.40	.52	.52
Educational Age	.48	.51	.53
Educational Quotient	43	.67	.67

Table 3 is very similar to Table 2, except that in this case the relationships shown are those between the *Kuhlmann-Anderson Intelligence Test* for Grade 1, and achievement measured at the beginning of Grade 4 The *Advanced Examination, Form W* of the *New Stanford Achievement Test* was used in this instance. The intelligence test scores are the same but the achievement test scores were obtained approximately six months later (after a summer vacation) and were obtained from a somewhat more advanced examination

None of the coefficients in Table 3 are significantly different from the corresponding coefficients of Table 2. The generalizations regarding Table 3 are the same as those drawn from Table 2 Excluding those coefficients involving E Q., the range of the coefficients of correlation is from 30 to .56, with a median r of 48, which is higher by only one σ_r of .04 than the median r of .44 in Table 2 In this instance it is again evident that the E Q is the most predictable measure of educational achievement. The reduction in errors of prediction due to the magnitude of r is practically identical with the corresponding reductions in Grade 3. The M A , I Q., and Pc.Av are about equally effective for prediction There seems to be no superiority of the Pc.Av over I.Q as a predictive measure

A study of Table 3 shows a higher correlation between all indices of intelligence (M.A., I.Q., and Pc.Av.) and arithmetic reasoning, than between the same indices of intelligence and any part of the reading tests. The correlation between the indices of intelligence and arithmetic average is higher than the correlations between the same indices of intelligence and reading average. The most noticeable shift in prediction is that of the arithmetic scores

Summary

The most predictable measure obtained from the *New Stanford Achievement Test, Primary Examination* (Grade 3) and from the *Advanced Examination* (Grade 4) when predictions are made from the *Kuhlmann-Anderson Intelligence Test* for Grade 1, seems to be the E.Q. ($r = .67$). Since E.Q. does not give any indication of the actual level of achievement, but only of achievement compared with C.A., this is significant only as an index of educational brightness The Pc.Av in this instance does not seem to have any advantage (for predictive purposes) over the I.Q. or M.A scores obtained from the Kuhlmann-Anderson Intelligence Test. All three indices of intelligence (M.A., I Q., and Pc.Av) are about equally effective for prediction.

Coefficients of correlation between the *Kuhlmann-Anderson* measures in Grade 1 and educational achievement in Grade 3 range from .32 to .53, and between the same test (*Kuhlmann-Anderson Intelligence Test, Grade 1*) and educational achievement in Grade 4 from .30 to .56. These low correlations indicate that long-range predictions of educational achievement based on only one group intelligence test in the first grade are highly questionable.

REFERENCES

1 Brown, M E. "Measuring Mental Ability in the Intermediate Grades of the Elementary School" *School and Society*, XXXV (1932), 323–324

2. Brown, A. W. and Lind, C. "School Achievement in Relation to Mental Age" *Journal of Educational Psychology*, XXII (1931), 561–576.

3 Buzley, D E "A Study of Test Results at the Third and Fifth Grade Levels" *Psychological Clinic*, XX (1931), 1–29

4 Cattell, P "The Heinis Personal Constant as a Substitute for the I.Q." *Journal of Educational Psychology*, XXIV (1933), 221–228

5 Durrell, D D "The Influence of Reading Ability on Intelligence Measures" *Journal of Educational Psychology*, XXIV (1933), 412–416.

6. Easley, H "One of the Limits of Predicting Scholastic Success" *Journal of Experimental Education*, I (1933), 272–276.

7. English, H D "The Predictive Value of Intelligence Tests" *School and Society*, XXVI (1927), 783–799

8 Erffmeyer, C. A. "Intelligence Tests as an Aid in the Diagnosis of Academic Maladies" *School and Society*, XX (1924), 307–320.

9 Gates, A. I "The Correlations of Achievement in School Subjects with Intelligence Tests." *Journal of Educational Psychology*, XIII (1922), 277–285.

10 Gates, A. I. "The Unreliability of M.A. and I.Q. Based on Group Tests of General Mental Ability." *Journal of Applied Psychology*, VII (1923), 93–100

11 Guilford, J P *Psychometric Methods* New York McGraw-Hill, 1936. Pp xi + 566.

12 Hawthorne, J. W. "The Effect of Improvement in Reading Ability on Intelligence Test Scores" *Journal of Educational Psychology*, XXVI (1935), 41–51

13. Hilden, A H "A Comparative Study of the Intelligence Quotient and the Heinis Personal Quotient" *Journal of Applied Psychology*, XVII (1933), 355–375.

14 Kelley, T L., Ruch, G M and Terman, L. M. *Guide for Interpreting the New Stanford Achievement Test.* Yonkers World Book, 1929 Pp 1–16.

15 Klein, A "Intelligence Compared with Achievement." *High Points Bulletin*, XII (1930), 3–5

16. Kuhlmann, F. "The Kuhlmann-Anderson Intelligence Tests Compared with Seven Others" *Journal of Applied Psychology*, XII (1928), 545–594.

17 Kuhlmann, F and Anderson, R *Kuhlmann-Anderson Intelligence Instruction Manual*, IV. Philadelphia: Educational Test Bureau, 1933 Pp. iv + 125.

18. Line, W. and Glen, J. S. "Some Relationships Between Intelligence and Achievement in the Public School." *Journal of Educational Research*, XXVIII (1935), 582–599
19. McCall, Wm. *Measurement* New York Macmillan, 1939. Pp xv+535
20. Mitchell, A. C. "Prognostic Value of Intelligence Tests" *Journal of Educational Research*, XXVIII (1935), 577–581
21. Riley, G. L. "A Comparison of the Personal Constant and Intelligence Quotient" *Psychological Clinic*, XVIII (1930), 26–65
22. St. John, C. S. *Educational Achievement in Relation to Intelligence* Cambridge, Massachusetts Harvard Press, (1930) Pp xiv+208.

MEASUREMENT ABSTRACTS*

Arthur, Grace. "A Non Verbal Test of Logical Thinking." *Journal of Consulting Psychology*, VIII (1944), 33–34

The author has formulated a non-verbal test of logical thinking, similar in purpose to the *Kohs Block Design Test*, but employing designs to be reproduced with plain and stenciled cards in various colors The problems presented by the designs were of three kinds, form, color, and sequence Tentative norms established on the basis of results obtained from 300 subjects indicate an increase in average score from one age group to the next. *Catherine Anne McNelly*

Bergmann, Gustav and Spence, Kenneth W "The Logic of Psychophysical Measurement" *Psychological Review*, LI (1944), 1-24

A methodological analysis of some of the problems of psychophysical measurement and of other aspects of measurement in psychology is presented from the standpoint of scientific empiricism After a discussion of the methodological frame of reference, in which the necessity for operational definitions from a physicalistic basis is stressed, and a review of certain principles of physical measurement, an analysis of psychophysical measurement is given. It is concluded that not only should the use of various terms applicable to physical measurement be discouraged in psychophysical measurement, but that measurement in psychophysics should be set up as a 'technique in its own right." *Lorraine Bouthilet*

Brown, Fred. "An Experimental Study of the Validity and Reliability of the Brown Personality Inventory for Children." *Journal of Psychology*, XVII (1944), 75–89

The *Brown Personality Inventory for Children* was administered to 77 clinically diagnosed neurotic boys and 200 normal boys between the ages of 8 and 15 and in Grades 4–9, inclusive, in order to determine whether the inventory would differentiate reliably and consistently between maladjusted and normal children. Highly significant differences between the two groups were found in each of the five categories of the instrument. These data were supplemented by two additional experiments with the personality inventory, which resulted in high test-retest correlations and consistently high reliabilities for the stability of individual items. *Catherine Anne McNelly*

Brown, Fred "Comparative Study of the Intelligence of Jewish and Scandinavian Kindergarten Children" *Journal of Genetic Psychology*, LXIV (1944), 67–92.

Three hundred and twenty three (131 males, 192 females) second generation Scandinavian and 324 (178 males, 146 females) second generation Jewish kindergarten children in the Minneapolis Public Schools were tested on the 1916 *Revision of the Stanford-Binet* Rigorous control of age, sex, and socio-economic status in selection of the data was exercised Comparisons made between the performance of the two groups on general intelligence, basal age, and vocabulary show no significant differences However, there appears to be a difference when the various subtests are considered Variation among occupational levels is greater for each of the group studied than the variation within individual occupational levels, and the difference that appears between the two groups decreases as one passes from lower to upper occupational levels *Miriam D Roiman.*

---

* Edited by Forrest A. Kingsbury

Cattell, Raymond B "An Objective Test of Character-Temperament: II" *Journal of Social Psychology*, XIX (1944), 99–113.

This study encompasses three distinct researches the first dealing with a group of 60 school children, the second with 49 adult women, and the third with 40 students. The objective of the study was to determine how far a personality trait recognizable in everyday life situations can be made to express itself also in a miniature laboratory situation, objectively scorable and requiring no great complexity of apparatus The results indicate that the *Character-Temperament Test* so formulated has high consistency, is almost uncorrelated with intelligence, and shows no significant sex difference in mean performance *Catherine Anne McNelly*

Festinger, Leon. "A Statistical Test for Means of Samples from Skew Populations." *Psychometrika*, VIII (1943), 205–210.

This paper presents a test for determining significance of differences between means of samples which are drawn from positively skewed populations, more specifically, those having a Pearson Type III distribution function The quantity $2np\bar{x}/m$ (where p equals the mean squared divided by the variance and n is the number of cases in the sample), which distributes itself as Chi Square for $2np$ degrees of freedom, may be referred to the tables of Chi Square for testing hypotheses about the value of the true mean. For two independent samples, the larger mean divided by the smaller mean, which distributes itself as F for $2n_1p_1$ and $2n_2p_2$ degrees of freedom, may be referred to the F distribution tables for testing significance of difference between means. The test assumes that the range of possible scores is from zero to infinity When a lower theoretical score limit (c) exists which is not zero, the quantity (Mean $-c$) should be used instead of the mean in all calculations. (Courtesy *Psychometrika.*)

Findley, Warren G (Chairman of the Committee on Psychological Tests) "Psychological Tests and Their Uses." *Review of Educational Research*, XIV (1944), 1–27

This issue reviews the literature for the three years ending July, 1943 The following titles are included

1. Findley, Warren G "Brief Overview of the Period"
2. Cornell, Ethel L. "Current Construction and Evaluation of Intelligence Tests."
3. Freeman, Frank S. "Applications of Intelligence Tests."
4. Sells, Saul B "Measurement and Prediction of Special Abilities."
5. Traxler, Arthur E. "Current Construction and Evaluation of Personality and Character Tests"
6. Darley, John G. and Anderson, Gordon V "Applications of Personality and Character Measurement."
7. Symonds, Percival M., Krugman, Morris, and Albert, Kathryn. "Projective Methods in the Study of Personality"
8. Findley, Warren G. "Measurement of Psychoeducational Growth" *Lorraine Bouthilet*

Fleming, Virginia "A Study of the Subtests in the Revised Stanford-Binet L and M." *Journal of Genetic Psychology*, LXIV (1944), 3–36.

The author systematically investigates the problem of whether or not the subtests in the *Revised Stanford-Binet* tests were appropriately placed in reference to difficulty She submitted 210 Form L and 118 Form M *Stanford-Binets* to three methods of analysis: (1) calculation of percentage of successes on each subtest, (2) calculation of the critical ratio for the difference in percentage of the same mental age group passing each subtest, (3) a refinement of the second method, where cases were selected only if they had taken every subtest within a year level Using the first method she found that there are several levels where the subtests are of unequal difficulty However, in comparing her results to those of Barber, who conducted a similar study, she found that they do not agree as to which are of unequal difficulty The last two more refined methods of analysis showed that in general subtests within the same level were of equal difficulty. *Miriam D. Roiman.*

Greenwood, Edward D., Snider, Hervon L. and Senti, Milton M "Correlation Between the Wechsler Mental Ability Scale, Form B, and Kent Emergency Test (E–G–Y) Administered to Army Personnel." *American Journal of Orthopsychiatry* XIV (1944) 171–173

Two hundred maladjusted army men were given the *Wechsler Mental Ability Scale, Form B,* and the *Kent Emergency Test, E–G–Y* As a result, a coefficient of correlation of .74 ± .02 was found between the Total Standard Score I.Q. of the *Wechsler Mental Ability Scale* and the *Kent Emergency Test* I.Q. Allowing for the abnormal group of men to whom the tests were given, this correlation was considered high The authors concluded that the *Kent Emergency Test* was a suitable intelligence test in situations which do not permit more extensive testing. *Catherine Anne McNelly*

Gulliksen, Harold "A Course in the Theory of Mental Tests." *Psychometrika*, VIII (1943), 223–245

An outline for a course in test theory is presented, together with a list of assignments, problems, and a bibliography The course has been given in the Psychology Department of the University of Chicago. The material is presented in outline form at the present time because of the increased need for training in test theory due to the increase in the use of psychological tests for classification of military personnel, and because much of the material in such a course must be selected from a wide array of articles in the literature. This material is presented in order that an organized body of material for instructional purposes may be readily available to those interested (Courtesy *Psychometrika.*)

Hunt, William A., Wittson, Cecil L. and Harris, Herbert I. "The Screen Test in Military Selection" *Psychological Review*, LI (1944), 37–46.

The authors compare the various paper-and-pencil psychological tests and the psychiatric interview as used in the pre-induction screening process They find the psychiatric interview method more flexible, more inclusive and easier to administer from the mechanical viewpoint, psychological test procedure is more economical both in manpower and time and is better standardized and more objective. As yet there is no final check as to which is the preferable procedure, but it is their belief that a good psychiatrist is a better screening instrument than a good test and a good test is better than a poor psychiatrist *Miriam D Roiman,*

Jurgensen, Clifford E. "A Nomograph for Rapid Determination of Medians." *Psychometrika*, VIII (1943), 265–269

Directions are given for constructing a very simple nomograph for computing medians, which is entered with information from the cumulative frequency distribution. It gives a linear interpolation within the class interval containing the median (Courtesy *Psychometrika.*)

Thornton, G R "The Significance of Rank Difference Coefficients of Correlation" *Psychometrika*, VIII (1943), 211–222

The coefficients of rank difference correlation that are barely significant at six different levels of significance are given for *N*'s of 2 to 30. Most of the values were obtained by translation of Olds' tables of probabilities for various values of Σd^2. Comparison of these data with those obtained by four other methods indicates that one method yields values more appropriate than those obtained from Olds' data for coefficients significant at the .01 level for *N*'s from 11 to 25. This method also provides a convenient means of obtaining approximate values of coefficients significant at the .01 level for *N*'s above 30. Need for caution in evaluating the significance of coefficients obtained from data involving tie rankings is indicated. The article concludes with recommendations as to choice of methods of determining the significance of rank difference coefficients. (Courtesy *Psychometrika.*)

Tinker, Miles A. "Speed, Power, and Level in the Revised Minnesota Paper Form Board Test." *Journal of Genetic Psychology*, LXIV (1944), 93–97

The *Revised Minnesota Paper Form Board Test* was administered to 103 university sophomores to analyze the relationship of the work limit method of measurement resulting in "speed" scores to the unlimited time method resulting in "level" scores and the relation of both of these to standard time method resulting in "power" scores. Speed and level scores were found to vary independently. A major proportion of the power score was accounted for by speed and level, with speed contributing relatively more to the power score than level. This study indicates only a slight correlation between intelligence and the *Revised Paper Form Board Test*. *Catharine Anne McNally*

Wherry, Robert J and Gaylord, Richard H. "The Concept of Test and Item Reliability in Relation to Factor Pattern." *Psychometrika*, VIII (1943), 247-264

It is shown that approaches other than the internal consistency method of estimating test reliability are either less satisfactory or lead to the same general results. The commonly attendant assumption of a single factor throughout the test items is challenged, however. The consideration of a test made up of K sub-tests each composed of a different orthogonal factor disclosed that the assumption of a single factor produced an erroneous estimate of reliability with a ratio of $(n-K)/(n-1)$ to the correct estimate. Special difficulties arising from this error in application of current techniques to short tests or to test batteries are discussed. Application of this same multi-factor concept to item-analysis discloses similar difficulties in that field. The item test coefficient approaches $\sqrt{1/K}$ as an upper limit rather than 1.00 and approaches $\sqrt{1/n}$ as a lower limit rather than .00. This latter finding accounts for an over-estimation error in the Kuder-Richardson formula (8). A new method of isolating sub tests based upon the item-test coefficient is proposed and tentatively outlined. Either this new method or a complete factor analysis is regarded as the only proper approach to the problem of test reliability, and the *item-sub-test* coefficient is similarly recommended as the proper approach for item analysis. (Courtesy *Psychometrika*.)

Lt. Comdr C. L. Wittson, U.S.N.R., Lt. Comdr W. A. Hunt, U.S.N.R., Lt. (JG) H. J. Older, U.S.N.R. "The Use of the Multiple Choice Group Rorschach Test in Military Screening." *Journal of Psychology*, XVII (1944), 91–94.

The Harrower-Erickson *Multiple Choice Group Rorschach Test* was tried out on a sample population, consisting of three groups, at the U. S. Naval Training Station, Newport, R. I. The test was unsuccessful in avoiding "false positives," picking up 44 per cent of the group of 417 normal subjects. It was unsatisfactory in picking out true positives, culling only 59 per cent of a group of 235 men "discharged as unfit for Naval service for neuropsychiatric reasons." Furthermore, of a group of 181 subjects previously "admitted to the observation ward for careful study but finally adjudged fit for service," the test picked up 59 per cent as belonging to the abnormal group. The authors found the test "unsuitable in its present stage of development for military selection." *Ralph J. Slattery*

NEWS NOTES*

An institute on student personnel work was held on the Los Angeles Campus of the University of California during the week beginning July 24, in connection with the 1944 Summer Session.

The institute was designed to help colleges and universities of the western states in the evaluation and development of student personnel services. It was planned in collaboration with Western Personnel Service, itself a cooperative association of western colleges and universities formed to work together on student personnel problems. The Academic Council of Western Personnel Service, under the chairmanship of Dean Karl Onthank of the University of Oregon, assisted Winifred Hausam, Director, and Helen Fisk, Associate Director, in the preparation of the program.

Leader of the institute was Dr E. G. Williamson, Dean of Students, University of Minnesota, President of the American College Personnel Association, Chairman of the Student Personnel Committee of the American Council on Education. During the war, Dr Williamson has been chairman of the Advisory Committee to the United States Armed Forces Institute; chairman of the Committee on Training of the Commission on Vocational Counseling of Veterans, War Manpower Commission, and consultant to the Adjutant General's Department concerning counseling of soldiers as part of the demobilization program.

Lt. Hugh M. Bell, A.G.D., is stationed at the Ninth Service Command Special Training Center, Camp McQuaide, California. The work of psychologists at the Special Training Center is described in an article by Lt. Bell and Lt. Altus in the *Psychological Bulletin*, March, 1944.

Francis F. Bradshaw is Dean of the College for War Training at the University of North Carolina, Chapel Hill, North Carolina.

Lucile B. Brown, formerly on the personnel staff at Northwestern University and later with Sears, Roebuck and Company, is overseas with the American Red Cross. For a time Mrs. Brown was in England and then was sent to North Africa. Her address is American Red Cross, A.P.O. 763, in care of Postmaster, New York City.

R. K. Compton, formerly Dean of the Division of General Science and Chairman of the Department of Psychology at South Dakota State College, is now Personnel Director, Hastings Manufacturing Co., Hastings, Michigan.

The new address of Lt. Wilbur S. Gregory, Guidance Consultant and Instructor in Psychology at the University of Nebraska, is Research Division, A.A.F. Instructors School (Flexible Gunnery), Laredo Army Air Field, Laredo, Texas.

Elias Lyman, Chairman of the Board of Personnel Administration at Northwestern University, resigned on May first and has returned to his old home at Lincoln, Vermont. F. George Seulberger, Professor of Cooperative Education and Chairman of the Department of Industrial Relations in the Northwestern Technological Institute, has been appointed Dean of Students, replacing Mr. Lyman.

---

* News items concerning members of the American College Personnel Association should be sent to Grace E. Munson, Northwestern University, Evanston, Illinois.

Lt. James A. McCharack, U.S.N.R., is on leave as Director of Personnel and Professor of Psychology at Brothers College, Drew University, and is stationed at State College, Pennsylvania.

Lt. (j.g.) Dewey B. Stuit, U.S.N.R., Associate Professor of Psychology at the University of Iowa (on leave), is Executive Officer, Naval Unit, Stevens Institute of Technology.

Frances O. Triggs, previously on the staff of the Personnel Bureau, University of Illinois, is now with the Social Security Board, Washington, D. C.

Ensign A. C. Van Dusen, U.S.N.R., formerly active in personnel work at the University of Florida, is now stationed at the Aerial Free Gunnery Central Standardization Committee, N.A.O.T.C., Naval Air Station, Jacksonville, Florida.

Mrs. Ada S. Westover, who has been Assistant Dean of Women at the University of Nebraska for the past ten years, has resigned to enter the field of medical social work in Cleveland, Ohio. Mrs. Westover had charge of the part-time employment of women students in addition to her work as Counselor.

Lt. C. Gilbert Wrenn, U.S.N.R., is acting as Secretary of the New Military Section of the American Association for Applied Psychology. Lt. T. Ernest Newland, U.S.N.R., also a member of the Section, attended the recent meeting held in Washington, D. C.

ANNOUNCEMENT

Following an almost unanimous vote by the membership to the effect that the nominating ballot should be considered as the final voting ballot, the following American College Personnel Association officers were designated for the year 1944-45:

President: E. G. Williamson,* Dean of Students, University of Minnesota, Minneapolis, Minnesota

Vice-Pres.: D. D. Feder, Bureau of Naval Personnel, Washington, D. C.

Secretary: Thelma Mills,† Director of Student Affairs for Women, University of Missouri, Columbia, Missouri

Treasurer: W. W. Blaesser,* Administrative Secretary of the Personnel Council, University of Wisconsin, Madison 6, Wisconsin

Members-at-Large of the Executive Council:

- J. L. Bergstresser, Dean of Students, College of the City of New York.
- A. J. Brumbaugh, Dean of Students, University of Chicago, Chicago, Illinois.
- Helen G. Fisk, Associate Director, Western Personnel Service, Pasadena, California
- Robert Hoppock, Professor of Education, New York University, Washington Square, New York, New York.
- Esther Lloyd-Jones, Professor of Education, Teachers College, Columbia University, New York, New York.

---

* Serving the second year of a two-year term.

† The secretary is elected for a two-year term (1944-46).

A TECHNIQUE FOR SCALE ANALYSIS

WARD H GOODENOUGH, Sgt, A.U.S

IN the past two years a considerable amount of time has been devoted to scale analysis by the Research Branch, Information and Education Division (formerly Morale Services Division) of the War Department, where the use of scales has been important in attitude and opinion research.

The aim of this paper is to make available to others working with scales one of the techniques for scale analysis developed in the Research Branch The theory involved in this type of analysis has been outlined elsewhere[1] and is not directly dealt with here

Four specific techniques have been employed by the Research Branch. All produce essentially the same results, differing only in the mechanics involved All have limitations either of a rigorous or practical nature

The first of these is the *least squares* method (Guttman, Louis, "The Quantification of a Class of Attributes" *The Prediction of Personal Adjustment*, Social Science Research Council Bulletin No 48, 1941). This method is too laborious to be readily usable when one is dealing with more than a few items and categories

A second method, the one generally used by the Research Branch, is to plot data on a *Scalogram Board*, invented for the purpose by Dr Louis Guttman This is an easily used system, especially if the sample is confined to 100 cases or less, which is usually adequate; but it is less rigorous, since manipulations of the data are based largely on inspection. The board has the further difficulty that it is somewhat expensive to construct.

---

[1] Guttman, Louis. "A Basis for Scaling Qualitative Data" *American Sociological Review*, IX (1944), 139-150.

A third method, the *trial-scoring and graphic technique*, has been used in the work of the Department of Sociology and Anthropology at Cornell University as well as in the Research Branch. It has the advantage that it does not require a board, but is less flexible

Finally, the *tabulation technique*, the one to be outlined here, was devised to enable scale analysis without a board, at the same time preserving its flexibility. It is also more rigorous than the Scalogram Board method and can be used with any size sample. Its limitation is that, for practical use, the relative rank order of categories for each item must be assumed in advance, which is not necessary with either the least squares or Scalogram Board technique The technique is easy to learn and as a process reveals clearly what is operationally involved in scale analysis.

According to Guttman's definition,[2] the multivariate distribution of a set of qualitative items forms a scale for a population if the following conditions are satisfied

(a) the items have sameness of content (that is, they form a universe of content),

(b) each item is a simple function of scores derived from the distribution.

Condition (a) is mainly determined by the nature of the problem an investigator is interested in. For example, suppose we are interested in students' expressions of adjustment to life at a particular college. For this purpose we might ask such questions as.

(1) How satisfied are you with college life?

(2) Do you think that what you are learning is worth while or not?

(3) What sort of a time are you having at college?

We ask these questions because they are a sample from the universe of content which is the object of our investigation Having defined the universe (in this case, adjustment to college life) and having collected a sample of data, we want to know whether or not the responses we have obtained behave as simple functions of a single variable If so, this will be evidence that the universe is a scale for the population, and we can derive meaningful relative adjustment scores based on the responses to the items

---

[2] *Ibid.*

This brings us to condition (b) of a scale. Can we show that each item is or is not a *simple* function of scores derived from the distribution of the items? The procedures to be discussed here refer mainly to this problem, the practical one of outlining a specific technique whereby it is possible with relative ease and precision to determine whether or not a set of items satisfy the formal condition (b) of a scale.

The question might arise as to why this is a problem at all If a set of items formed a perfect scale, this would be easily established by any of the several techniques The problem arises *when the scale-relationship of a set of items is obscured by the items' categories* This requires the *combining of categories*, which manipulation creates the main technical problem, as will be seen in the ensuing discussion

For purposes of illustration we shall concern ourselves with the three hypothetical questions on student adjustment to college life. We shall assume that each question has three categories, which we have assigned them on the questionnaire.

(1) How satisfied are you with college life?
   (a) Very satisfied
   (b) Neither satisfied nor dissatisfied
   (c) Very dissatisfied

(2) Do you think that what you are learning is worth while or not?
   (a) It is worth while
   (b) Undecided
   (c) It is not worth while

(3) What sort of a time are you having at college?
   (a) I am having a good time
   (b) I am having a fair time
   (c) I am having a miserable time

First, then, we must see what the conditions are under which each item can be said to be a simple function of scores derived from the distribution of the items To do this, suppose the responses to our three questions gave us the following marginal frequencies:

		Category			
		a.	*b.*	*c.*	*Total*
Item	1	25	20	55	100
	2.	20	60	20	100
	3.	40	30	30	100

We can express these marginal frequencies in bar chart form (Fig. 1).

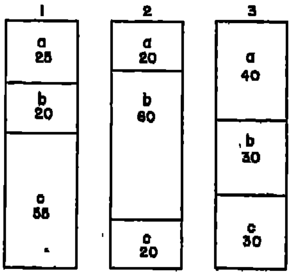
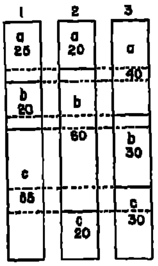


Figure 1

For the items to form a scale, that is, to be simple functions of scores derived from the items, all those respondents who make the positive response (a) to item 2 must also make positive responses to items 1 and 3, as Guttman has shown (*ibid*). This group of respondents would constitute a *scale type* In like manner, the other scale types can be read from the bar chart (*provided*, the items form a scale) simply by connecting the bars with dash lines (Fig. 2).

It is clear then, that, given the marginal frequencies to a set of items and the relative position of the categories within each item, it is possible to ascertain which response combinations, from among all those *possible* for the items, constitute the ideal scale types for those marginal frequencies.[3] But we still do not know whether or not the items actually scale We are faced with the problem of determining how close to the ideal the obtained response combinations and their frequencies actually come.



Frequency	Scale Type	Response Combination
20	1	a-a-a
5	2	a-b-a
15	3	b-b-a
5	4	b-b-b
25	5	c-b-b
10	6	c-b-c
20	7	c-c-c

Figure 2

If a tally of the response combinations obtained showed all to coincide with the ideal scale type combinations and their frequencies as determined from the marginal frequencies, the responses would form a perfect scale and there would be no further problem of procedure. In practice, however, such perfect conformity is almost never encountered. The response patterns obtained always vary to a greater or lesser degree from the ideal scale established for the marginal frequencies

The problem, therefore, is to determine to what extent the response patterns obtained approximate the ideal scale; to what extent the response patterns obtained may deviate from the ideal scale and still permit of analysis in scale terms, and what the procedural implications are when the response patterns deviate beyond such an established point of tolerance from the initial ideal scale

Guttman has set the point of tolerance for deviation at 15%. In other words, at least 85% of the total number of responses must fall within the scale pattern, so that it is possible to reproduce at least 85% correctly all the responses of all the respondents from the scale scores.

If there is more than 15% deviation from the initial ideal scale, we have evidence that there is more than one factor or variable involved This allows us two inferences. One is that the extraneous factors are of minor importance and mainly affect only the categories, in which case they can be partialled out by combining categories. The other inference is that there are major factors affecting the items themselves, in which case category combination will not partial them out and no scale can be established.

To illustrate what is involved where the first inference obtains, let us suppose that a count of all the possible response combinations for our three items on adjustment to college life showed the following results (When large samples are involved, this can easily be done on the counting sorter machine or with tabulating equipment if the data are put up on I.B.M. punch cards. Lacking such equipment, one may conveniently determine the nature of the scale relationship by working with a sub-sample of 100 respondents.)

Here we have deviation from the ideal scale well beyond the 15% point of tolerance. Our problem now is to examine the nature of the deviation.

A glance at the six response combinations which deviate from the ideal scale types shows that four of them have one thing in common: the response (c) to item 2. If we consider categories (c) and (b) on item 2 to be equivalent responses from the standpoint of the scale relationship of item 2 to the other items, then 15 of the deviant responses coincide with scale types: b–c–a is equivalent to the scale type b–b–a, b–c–b to b–b–b, and c–c–b to c–b–b Furthermore, the scale types c–c–c and c–b–c are equivalent and the sum of their obtained frequencies is equal to the sum of their ideal frequencies.

---

[3] It is possible to determine the maximum number of scale types though not the combinations of responses which constitute them by using Guttman's formula Maximum number of scale types equals sum of categories in all items plus one minus number of items ("The Desire of Enlisted Men for Post-War Full-Time Schooling An Example of a Scale." Mimeographed Paper, Research Branch, Morale Services Division, ASF, 1943)

Ideal Scale Type Combinations *Item* 1	2	3	*Ideal Frequency*	*Obtained Frequency*
a	a	a	20	15
a	b	a	5	5
b	b	a	15	5
b	b	b	5	5
c	b	b	25	10
c	b	c	10	30
c	c	c	20	0
Total			100	70

Other Combinations Obtained *Item* 1	2	3	*Obtained Frequency*
a	a	b	5
c	b	a	5
b	c	a	5
b	c	b	5
c	c	a	5
c	c	b	5
			30

Therefore, in order to reduce deviation from the ideal types and frequencies to less than 15%, it is possible to combine categories of any item. Each combination eliminates a discriminating scale type, however, and a minimum of combination is therefore desirable.[4] It is to be noted that since a combination involves the assumption that two categories are equivalent in the scale, it is ordinarily possible to combine only categories which are adjacent to each other in the category hierarchy of an item. In our present example we can combine positive (a) with neutral (b) or neutral (b) with negative (c), but not positive (a) with negative (c) without including neutral (b), which would eliminate the item from the scale as having no discriminatory value.

To facilitate category combination it is generally desirable to set up a distribution table such as the one below (Fig 3). For purposes of illustration we shall show in the table all the possible combinations of response, which ones are the ideal scale types, and where the obtained frequencies fall Using our three items on adjustment to college life, we get the following:

a									b									c									Item 1.
a			b			c			a			b			c			a			b			c			Item 2.
a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	Item 3
20			5									15	5									25	10			20	Ideal Freq.
15	5		5									5	5		5	5					5	10	30	5	5		Obtained Freq.
1			2									3	4									5	6			7	Scale Type

Figure 3

When we combine categories (b) and (c) on item 2, as we have already ascertained to be desirable, our table looks as follows (Fig. 4):

a						b						c						Item 1.
a			b&c			a			b&c			a			b&c			Item 2
a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	Item 3.
20			5						15	5						25	30	Ideal Freq.
15	5		5						10	10					10	15	30	Obtained Freq.
1			2						3	4						5	6 7	Scale Type

Figure 4

In order to reduce the deviant frequency 10 for the response combination c–b&c–a (Fig. 4) we can combine either (b) and (c) on item 1 (in which case b–b&c–a coincides with c–b&c–a) or (a) and (b) on item 3. The latter will also reduce the deviation involved in the combination a–a–b. Thus, by the fewest possible combinations of category, we obtain the maximum possible reduction in deviation and emerge with the following table (Fig. 5):

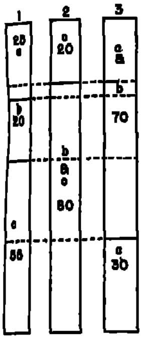
---

[4] In accordance with Guttman's formula in footnote 3.

a				b				c				Item 1.
a		b&c		a		b&c		a		b&c		Item 2.
a&b	c	a&b	c	a&b	c	a&b	c	a&b	c	a&b	c	Item 3.
20		5				20				25	30	Ideal Freq.
20		5				20				25	30	Obtained Freq.
1		2				3&4				5	6&7	Scale Type

Figure 5

The scale that emerges from the data actually obtained for our three items can now be presented in bar chart form (Fig. 6).



Frequency Ideal & Obtained	Scale Type	Response Combination
20	1	a a a&b
5	2	a b&c a&b
20	3&4	b b&c c&b
25	5	c b&c a&b
30	6&7	c b&c c
100		

Figure 6

We can now say that all those respondents who are included in response (a) of Item 2 are also included in (a) of item 1 and

(a&b) of item 3, and so on down the scale. It is possible to reproduce the responses of any respondent from whatever set of scores one may wish to use to designate the five scale types.

In other words, we can now discriminate five degrees of what we are calling adjustment to college life on the basis of our three items

1 Those who a) are very satisfied with college life
b) feel that what they are learning is worth while
c) have a good or fair time

2. Those who a) are very satisfied with college life
b) feel that what they are learning is not worth while or are undecided
c) have a good or fair time

3 Those who a) are neither satisfied nor dissatisfied
b) feel that what they are learning is not worth while or are undecided
c) have a good or fair time

4. Those who a) are very dissatisfied with college life
b) feel that what they are learning is not worth while or are undecided
c) have a good or fair time

5. Those who a) are very dissatisfied with college life
b) feel that what they are learning is not worth while or are undecided
c) have a miserable time

For purposes of illustration, all deviation was reducible in our example. In practice, however, complete reducibility of deviation is rarely possible without combining all categories of all items, thereby resulting in no discrimination at all.

When the amount of deviation is sufficiently reduced to make further combination of categories impractical, there is the problem of how to score the few remaining deviants (the deviation presumably having been reducible to 15% or less). Suppose we had had such a response combination as a–a–c on our

three items by the time we had done all practical category combination (Fig. 5). By scoring a–a–c with a–a–a&b we would have only one deviant response (c) on item 3. Any other scoring, as, for example, with c–b&c–c, would leave at least two deviant responses (a) on item 1 and (a) on item 2 Maximum reproducibility of response from the scale score is, therefore, the principle one would ordinarily desire to follow in the scoring of deviants. (It occasionally happens that a deviant can be scored in several different ways and still permit of maximum reproducibility. The indeterminacy arises if the number of items is small and can be removed by using more items. Usually the indeterminacy is harmless, however, and it is sufficient to give such deviants the most neutral of the several best possible scale scores.)

When many items and/or categories are involved, it is obviously too cumbersome to work with such tables as those above (Figs. 3, 4, 5), where all the *possible* combinations of response are shown. It is not difficult, however, to work with similar tables allowing space only for the response combinations actually obtained By omitting from the table all those response combinations which have a zero frequency, the resulting abbreviated table is convenient. To illustrate, we could have set up the first table in our example with the three items on adjustment to college life (Fig. 3) in the manner shown below (Fig. 7).

a			b				c						Item 1.
a		b	b		c		b			c			Item 2.
a	b	a	a	b	a	b	a	b	c	a	b	c	Item 3.
20		5	15	5				25	10			20	Ideal Freq.
15	5	5	5	5	5	5	5	10	30	5	5	0	Obtained Freq.
1		2	3	4				5	6			7	Scale Type

Figure 7

In the course of our discussion we have seen that it is possible to determine from the marginal frequencies of a set of items what the scale types should theoretically be for those items and for the population responding to them. Then, by comparing these with the response combinations obtained and by combining categories, if necessary, we have observed how the existence of a scale can be established, how the remaining deviation can be determined, and what the principles are which govern the scoring of deviants

In conclusion, it is important to remember an assumption which has made possible the procedure outlined Everything that we did involved the premise that the hierarchy of category within each item was reasonably correct to begin with, even where we made combinations of category If the content of our categories had been such that no assumptions could have been made as to their relative positive or negative value under each item, we could not have set up an initial ideal scale for the items, a step on which our entire procedure was dependent.

From the practical standpoint, this limitation is not too important for work in the field of attitude and opinion, since the order of categories within items can usually be judged beforehand with sufficient accuracy. In some cases, however, it may be difficult to do this Any of the techniques for scale analysis can handle this problem of determining category order "blindly", but it will ordinarily involve far more labor than in the case (such as that outlined in this paper) where the correctness of *a priori* judgments is assumed. The tabulation technique of this paper is being extended to include this most general case where the category order is not known

A STUDY OF THE USE OF MECHANICAL APTITUDE TESTS IN THE SELECTION OF TRAINEES FOR MECHANICAL OCCUPATIONS

J W McDANIEL
Bakersfield Junior College
AND
WM. A REYNOLDS
Sacramento Air Service Command

WARTIME dislocation of the labor market in the direction of mechanical occupations emphasizes a need for selective screening of trainees Postwar return to trade, service, and professional occupations will demand still greater attention to applicants for retraining. If schools are to retain control over their programs, they must provide for increased vocational guidance for youth just preparing to enter the labor market and for adults seeking occupational readjustment. Increased guidance demands improved methods for determining potential occupational fitness Training programs now in the secondary schools give the schools a fine opportunity for developing selection and guidance procedures useful during wartime and indispensable in vocational relocation after the war.

The ideal test of job fitness is a tryout experience under job conditions and under expert observation. The practical limitations on the wide use of this selective device necessitates construction of substitutes Two general types of such substitutes have been devised One type of measuring instrument is a miniature work sample in which effort is made to duplicate the work situation under measurable conditions Another type of instrument attempts to analyze the measurable abstract elements of the job activities. This type may reduce the situation to elements that can be tapped in a language form or other visual pattern and results in pencil-and-paper tests of aptitude. A number of these tests are on the market, and some

of them claim high correspondence with actual work success For most of them very limited data are available showing their relationship to trainability of potential workers. The great need is for specific local evidence of usefulness of such tests in selecting trainees for middle level occupations

This paper reports a study begun in the spring semester of 1942 at the Kern County Union High School and Bakersfield Junior College for the purpose of forming a test battery to predict success in mechanical training courses The study was based on the results of administering three tests purporting to measure mechanical aptitude to 206 students, of whom 112 were in high-school classes and 94 in National Defense Training programs. Trainees in machine shop, aircraft engines, aircraft mechanics, wood shop, and welding were tested on the *Bennett Test of Mechanical Comprehension*, the *MacQuarrie Mechanical Ability Test*, and the *O'Rourke Test of Mechanical Ability, Junior Grade, Form C.*

The criterion of success used to validate the tests was the median rating on four traits given the trainee by the instructor. After careful instructions in the use of the mechanical aptitude rating scale the instructors rated the trainees on. (1) aptitude for learning as evidenced by ability to understand and apply instructions to the job, (2) speed and accuracy as evidenced by ability to acquire muscular and manipulative skills, (3) workmanship as evidenced by quality and precision of work done, and (4) interest-enthusiasm as evidenced by eagerness in getting at the job and staying with it. Ten-point scales were constructed and behavior descriptions placed underneath various points on the scales.

Results

Of the 206 students, 197 were males and 19 were females The females and those who had incomplete test scores were eliminated, leaving 147 males in the validating group The means and standard deviations of the various tests for this group are presented in Table 1. The data on the high-school students and National Defense trainees are presented separately. Although the National Defense group is superior on

the Bennett and O'Rourke tests, the critical ratios between the two groups do not exceed 2.7,[1] and for the purposes of this study the two groups may be considered to be comparable.

The intercorrelations of the subtests of the battery are shown in Table 2 Since the standard error of an r of zero for 147 cases is .082, it is seen that only forty-two of the intercorrelations are significantly greater than three times the standard error of a chance correlation However, an attempt was made to combine some of the subtests so that a significantly higher multiple-correlation coefficient would be obtained.

TABLE 1

The Means and Standard Deviations of the Subtests of the Mechanical Aptitude Test Battery for the Groups Used in the Validation Study

Test	High School (N = 80)		National Defense (N = 67)		Total (N = 147)	
	M	S.D.	M	S.D	M	S D
Bennett	40.25	8.80	44.02	8.12	41.97	8.70
MacQuarrie 1	38.31	8.79	39 51	7 74	38.86	8.35
MacQuarrie 2	38.31	6.90	40.58	7 70	39.35	7.36
MacQuarrie 3	20.20	3.00	20.10	3.37	20.15	3 17
MacQuarrie 4	39.69	12.54	40.21	14.96	39 93	13 70
MacQuarrie 5	21 13	8.04	22 92	9 16	21.94	8.61
MacQuarrie 6	13.50	6.67	15 15	6.34	14.34	6.59
MacQuarrie 7	20.75	4.44	20.32	6.16	20.55	5.55
MacQuarrie total	63.85	9 74	66.24	11.99	64.94	10.89
O'Rourke I	80.26	17.91	86.06	12.86	82.91	16.08
O'Rourke II	151.50	40.45	168.00	38.82	159.56	40.46
O'Rourke total	231.90	52.13	252 78	49.69	241.74	52.03
Criterion (ratings)	5.38	1.40	5 70	1.51	5.63	1 45

The first combination of subtests to be tried consisted of subtests B, M_1, M_5 and O_1, as these had higher coefficients with the criterion while at the same time having lower correlations with the other tests. The corrected[2] multiple correlation coefficient was found to be .35, and the combination was rejected as being too low in predictive power

The second battery tried was the combination of all sub-

tests, excluding the totals of the MacQuarrie[3] and O'Rourke tests. The corrected multiple-R was .43 When the predicted criterion scores were correlated with the obtained criterion scores the correlation was found to be slightly, but not significantly, higher, or 45 [4] Because of the labor in forecasting from ten tests and because the efficiency of prediction is but 10 7% better than chance for this battery, a combination was sought which would contain fewer tests but would give as high an efficiency of prediction.

TABLE 2

Intercorrelations of the Subtests of the Mechanical Aptitude Battery for 147 Male Trainees

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	O_1	O_2	O_3	M	C
B	.11	13	16	43	.39	56	.21	.37	.56	.55	48	.34
M_1		.18	.32	17	13	.20	.23	-.02	.05	.03	.51	.22
M_2			.31	.22	.28	15	.34	-.08	.05	.04	47	-.17
M_3				18	41	.22	.30	05	.09	12	.62	.22
M_4					.55	.50	.41	.30	.29	.34	73	.21
M_5						51	.34	-.01	31	34	78	10
M_6							.49	.39	.43	.51	.67	.22
M_7								14	.14	.06	.65	.13
O_1									.52	81	.28	.21
O_2										.96	.33	.28
O_3											.51	.26
M_t												.34

S D.$_{r=0}$ = .082

Legend
B = Bennett Test of Mechanical Comprehension
M_1 = MacQuarrie Tracing
M_2 = MacQuarrie Tapping
M_3 = MacQuarrie Dotting
M_4 = MacQuarrie Copying
M_5 = MacQuarrie Location
M_6 = MacQuarrie Blocks
M_7 = MacQuarrie Pursuit
M_t = MacQuarrie total score
O_1 = O'Rourke Mech. Apt., Part I
O_2 = O'Rourke, Part II
O_3 = O'Rourke, total score
C = Criterion (instructors' ratings)

The third trial battery consisted of the seven tests B, M_1, M_2, M_3, M_4, M_5 and O_1. The corrected multiple-correlation coefficient was found to be .45, and when computed by correlating predicted with obtained criterion scores, 47. The effi-

---

[1] A critical ratio (D./S.D.$_{diff}$) of 3 0 is considered to indicate an almost certain difference between two groups.

[2] Ezekiel, M. *Methods of Correlation Analysis* John Wiley & Sons, Inc., New York, 1930, p. 177 This correction is for a limited number of cases and for the number of variables, and estimates the probable correlation for the universe from which the data were drawn.

[3] The total of the MacQuarrie Test never entered into any other combination. Since its correlation with the criterion is lower than some of its subtests, the addition to a battery merely attenuates the multiple-correlation coefficient.

[4] A multiple-R computed from the Beta weights should equal exactly the R obtained by correlating the predicted scores with the obtained scores. The differences were due to (1) errors in the facilitating tables which were used in making the predictions, when the four-place numbers were rounded off to two places, and (2) errors due to grouping not compensated for entirely by Sheppard's correction.

TABLE 3

The Multiple Correlation Coefficients of the Four Trial Batteries and their Efficiencies in Prediction

Tests in battery	R	E
[illegible]	33	.056
[illegible]	45	.107
[illegible]	.47	117
[illegible]	.41	.088

ciency of prediction for this battery was slightly better, being 11 7% better than chance.

A fourth battery was tried, consisting of the five tests B, M_1, M_2, M_3, and M_4, and the corrected multiple-R was 42. The correlation between the predicted and criterion scores showed a slight drop to 41 The efficiency of prediction for this battery was thus only 8 79% better than chance.

The coefficients of multiple-correlation for the four trial batteries are summarized in Table 3. The efficiency of prediction for each multiple-R is shown in the last right-hand column

Since the multiple-correlation coefficient of the third battery indicates that it is the best of the obtained combinations, the predicted scores were correlated with the obtained criterion scores for each separate class of mechanical work.[a] The data were summarized in Table 4

From Table 4 it may be noted that the multiple-R varies widely from class to class The coefficient for the high-school Machine Shop class, $R = .60$, is sufficiently high for forecasting purposes, while for the National Defense A/C Machine Shop class, $R = -.42$, it is useless An examination of the means and standard deviations of the classes shows that one instructor, Ro, rated much more severely than the others, and the negative correlation indicates a different standard of scoring in that class. It is possible that the results of the study would have been affected in the direction of greater forecasting efficiency if the data for this class had been omitted

TABLE 4

The Correlations between Obtained and Predicted Criterion Scores for Each Mechanical Class

Class	Group	Instructor	N	Observed criterion		Predicted criterion		R
				M	S D.	M	S.D	
Machine Shop	High school	P	36	5.97	1 41	5 73	.55	.60
Machine Shop	Nat. Def.	P	15	6.73	1.15	6.26	.69	38
A/C Mechanics	Nat. Def.	R	24	5.92	1 19	5 73	47	48
Woodshop	High school	Sc	23	5 13	.95	5.50	44	.37
Welding	High school	St	21	5.38	1.59	5.63	.88	.27
Welding	Nat. Def	St	14	5.79	1.80	5 19	.90	.42
A/C Machine	Nat. Def	Ro	14	4.14	43	5.33	.39	−.42
All classes	H S. & N D		147	5.63	1.45	5.64	.67	.47

---

[a] The raw-score regression equation for the third battery is .013 (B) + .028 (M_1) − .087 (M_3) + .112 (M_6) + .019 (M_7) − .022 (M_8) + .007 (O_1) + 2.540

The small number of cases in the classes, however, should be noted as a factor introducing unreliability in the results shown in Table 4, as strict comparability between the instructors' ratings could not be obtained by transmuting them to a scale with a common mean and sigma Perhaps more important than this is the fact that the ratings were taken to the nearest whole number on a ten-point scale, and were not sufficiently fine for meaningful transmutation.

TABLE 5

Regression Weights of the Subtests of the Mechanical Aptitude Test Battery

Test	Symbol	Weight
[illegible]	B	.0788
[illegible]	M_1	.1580
[illegible]	M_3	−.2878
[illegible]	M_6	.2439
[illegible]	M_7	.1796
[illegible]	M_8	−.1306
[illegible]	O_1	.2027

Further evidence for the validity of the forecasting equation was obtained by predicting the scores of the nineteen women not included in the validating group. Fourteen of them were in a class on Aircraft Engines, under an instructor not listed in Table 4, and five women were in the National Defense A/C Machine Shop class listed in Table 4. The multiple-R for this group was .78 The coefficient is slightly greater than three times the standard error of an r of zero for 19 cases.

The regression coefficients (Beta weights) for the tests are given in Table 5.

It is evident from the relative magnitudes of the Beta weights that the MacQuarrie Tapping and Dotting Tests, and the O'Rourke General Information Test contribute the most to the forecasting efficiency of the battery

Summary and Conclusions

1 Three standardized tests of mechanical aptitude were administered to 147 high-school students and National Defense trainees in five mechanical training courses.

2. From the three major tests a combination of seven subtests was selected which had a multiple-correlation coefficient with the criterion of instructors' ratings of .47 It is probable that this value would have been raised if two criterion measures had been obtained and an estimate of the true correlation of the battery with the criterion had been made.

3 The forecasting efficiency of the battery differed for the five training courses, but its over-all value is significant in showing that prediction formulae need not be confined to but one type of training. Evidence, while not conclusive, was obtained that the formula would forecast a new group with an efficiency greater than if no correlation obtained.

4 Weightings of the subtests of the battery indicate that the MacQuarrie Tapping and Dotting Tests, with the O'Rourke General Mechanical Information Test, contribute most to the forecasting efficiency of the battery.

A VALIDATING STUDY OF THE OBERLIN VOCATIONAL INTEREST INQUIRY

L. D HARTSON
Oberlin College

IN THE YEAR 1927–28 W H. Brentlinger and the present author devised a *Vocational Interest Inquiry* for use in connection with the counselling service which was at that time one of the responsibilities of the Department of Psychology at Oberlin College. Inasmuch as a description of this *Inquiry* may be found in Fryer's *Measurement of Interests* (1), it will be unnecessary to reproduce it here. A statement may well be made, however, as to the hunch which motivated the construction of this particular form of inquiry. A rather detailed classification had just been completed of the occupations in which the Oberlin alumni were engaged (2) This experience suggested the desirability of a vocational inquiry from the results of which a man's interests could be classified in terms of the general type of function which appealed to him most strongly. A list of twenty functions was therefore formulated, quite *a priori*, and enough items selected to permit a paired comparison between each function and each of the nineteen others. The *Inquiry* involves, therefore, 190 pairs of comparisons and yields a score in each of the following twenty categories:

I. Creative Activities
II Research
III. Social Investigation
IV Diagnosis
V Compilation of Previously Discovered Facts
VI. Oral or Written Presentation of Reports
VII. Financial Speculation
VIII. Deeds of Prowess
IX. Formulation of Policies
X. Detailed Planning of Work
XI. Executive Management of Personnel
XII. Practical Construction Problems
XIII. Routine Activities
XIV. Evaluation of Conduct, Achievement or Product
XV Arbitration
XVI. Administration of Advice
XVII. Selling Ideas or Commodities
XVIII. Teaching
XIX. Offering Service or Aid
XX. Professional Entertainment

The principal use made of the *Inquiry* was as an aid in the counselling of students. With a change in the administrative setup, made in 1928, with reference to vocational counselling, use of the *Inquiry* was discontinued In order that some conclusion might be reached as to the validity of this form of vocational inquiry, a follow-up study was initiated in 1941 by E B Knauft and the author Copies of the blank were mailed to 178 men who were the original subjects. Responses were received from 121, or 68 per cent of the list It is unfortunate that the population is not large enough to furnish more definitive information concerning the characteristic patterns of interest for contrasting vocational groups. The number is large enough, however, to furnish some suggestions as to the consistency of individual interest patterns over the period of fourteen years—time enough to permit most of the men to become established in their professions

Validation of the Inquiry

1. *Extent of Agreement between the Patterns of Interest Indicated by the Inquiries of 1927 and 1941.* By ranking the functions on the basis of the frequency with which they were preferred, it is possible to compute rank difference correlations between the two sets of choices. For the total population of 121 the consistency of choices on the two occasions is indicated by an *r* of .717 ± .077 (see Table 1). The correlations for the different vocational groups range between .899, for the 12 men in the field of social science, to .282 for five accountants; for nine individuals who could not be fitted into any group classification, from .842 to .063 The table also reports the correlations for three special aggregates For the nine artists and musicians in this sample, consistency of interest is represented by a correlation of .889, for the 27 college teachers, by an *r* of .804, and for the 30 men in business and industry, by an *r* of .668.

2. *Agreement between the 1927 and 1941 Responses, as Indicated by the Most and Least Favored Functions.* Table 2 reports the three most favored and three least interesting functions, on the occasion of each of the *Inquiries*, for each subgroup This table shows that in the case of those groups for which the correlation between the two responses showed a correlation as high as .726, at least one of the functions was among the three most favored on both occasions, and with all but two of the groups this is true of two or three of the functions. When one turns to the least preferred functions, it is desirable to eliminate from consideration function XIII (Routine Activities), since that is almost universally unpopular No other function is commonly eschewed, but at least one of the other two functions which was least interesting on the first occasion is found at the bottom of the list in 1941 with all groups for whom the over-all correlations between the two series of choices is at least .648, as well as with three other vocational groups where the correlation is lower than that figure

TABLE 1

Correlations between Test and Re-Test Oberlin Vocational Interest Inquiry

Occupational group	N	*r*	P.E.
Social Scientists	12	.899	.030
Artists	4	.870	.038
College Language Teachers	8	.804	.056
Musicians	5	.746	.070
Librarians	2	.746	.070
Ministers	3	.726	.075
Physicians	8	.726	.075
Physical Scientists	6	.726	.075
Industrialists	7	.648	.092
Chemists	7	.618	.094
Biologists	5	.618	.098
Advertising Managers	2	.618	.098
Bankers	6	.558	.108
Retailers	9	.507	.117
Lawyers	7	.446	.126
Physical Education Instructors	3	.393	.133
Geologists	3	.393	.133
Secondary School Men	10	.292	.144
Accountants	5	.282	.145
Air Corps Instructor	1	.842	.046
Social Worker	1	.832	.048
Personnel Man	1	.558	.108
Actuary	1	.467	.123
Editor	1	.436	.128
Director, Social Planning	1	.385	.135
Criminal Court Psychologist	1	.344	.139
Editor	1	.292	.144
Agent, Transportation	1	.063	.157
Total population	121	.717	.077
Combined groups:			
Artists and Musicians	9	.889	.031
Business and Industry	30	.668	.087
College Teachers	27	.804	.056

TABLE 2

Functions of the Oberlin Vocational Interest Inquiry Most and Least Preferred by Each Occupational Group

	N	Most preferred		Least preferred	
		1927 test	1941 test	1927 test	1941 test
Social Scientists	12	2-4-14	4-9-2	13-8-20	13-8-20
Artists	4	1-2-10	2-1-18	13-17-15	7-8-13
College Language Teachers	8	1-2-14	1-2-14	13-7-12	7-8-17
Musicians	5	9-18-20	20-1-16	8-3-13	8-12-13
Librarians	2	9-15-17	2-6-9	12-13-8	12-13-8
Ministers	3	6-9-14	14-6-15	12-7-13	12-7-8
Physicians	8	19-2-4	4-16-19	13-5-7	13-6-7
Physical Scientists	6	2-1-12	2-4-1	17-16-15-5	13-17-20
Industrialists	7	8-4-11	1-9-2	5-13-7	7-13-20
Chemists	7	2-12-8	2-1-9	6-5-20	13-7-17
Biologists	5	1-8-4	1-4-9	7-6-13	13-7-17
Advertising Managers	2	7-9-11	9-11-16	13-12-6-18	12-20-19
Bankers	6	1-2-8-11	16-9-4	20-5-13	6-12-13-20
Retailers	9	8-2-1	9-4-16	13-17-20	12-13-5
Lawyers	7	1-8-20	15-16-18	13-5-10	13-7-12
Physical Education Instructors	3	19-8-1-2-4	1-10-9	13-6-5	13-6-20
Geologists	3	1-11-20	2-1-4	7-13-17-19	13-8-15-6
Secondary School Men	10	8-15-19	2-1-4	13-7-5	7-13-20
Accountants	5	4-13-5	14-2-11	6-10-15-17	8-6-13
Air Corps Instructor	1	1-4-8-9	1-2-4-11-14	5-13-6-15	20-15-6
Social Worker	1	18-11-19	15-11-18	5-6-2-10-13	5-6-10-19
Personnel Man	1	3-14-9	15-19-16	12-5-13-20	13-2-5-6-12
Actuary	1	1-7-2	2-9-1	15-6-13-17	8-15-7-20
Editor	1	5-4-20	12-4-5-9-11	7-16-18	7-15-13
Director, Social Planning	1	9-11-1-5	2-9-16-18	2-12-13	8-3-11-15-20
Criminal Court Psychologist	1	8-12-1	2-1-19	20-17-7	2-13-7
Editor	1	1-8-19	9-11-4	13-10-16-5	20-12-13
Agent, Transportation	1	8-19-12	2-4-1-5	3-5-6-7-9-13-14	7-13-8
Total population	121	1-2-4	1-4-2	13-17-12	13-7-12

3 *Agreement of the Profiles with Empirical Judgment as to the Functions Characterizing the Different Vocations.* Examination of Table 2 shows that the pattern of preferred functions shows marked contrasts when one vocation is compared with another, and that they conform almost without exception to common sense expectation Examination of the responses made in the 1941 *Inquiry* shows, e g., that the three functions chosen most frequently by the social scientists are: Diagnosis, Formulation of Policies, and Research In response to the initial *Inquiry*, this group showed a strong interest, also, in Evaluation of Conduct On the other hand, they displayed least interest in Routine Activities, Deeds of Prowess, and Professional Entertainment. The physicians show a distinctly different pattern Their initial preferences were for Diagnosis, Research, and Offering Service or Aid, but by the time they had begun the practice of medicine, the function of Administration of Advice had displaced Research The physical scientists showed a similar preference for Research and Creative Activities, but substituted Diagnosis for their initial choice of Practical Construction Problems. The group of college teachers who were working in the general area of the languages report unanimity in their first three choices on both occasions: Creative Activities, Research, and Evaluation of Achievement or Product. The types of activity most attractive to the clergymen are: Oral Presentation of Reports, Evaluation of Conduct, and either the Formulation of Policies or Arbitration (the settlement of disagreements). The final preferences of the lawyers differ widely from the undergraduate choices. They present a logical combination. Arbitration, the Administration of Advice, and Teaching.

On the other hand, there are instances in which the pattern chosen seems unrelated to the occupation. This is notably the case with the men in the field of secondary education, whose more recent interests lie in the field of Creative Activities, Research, and Diagnosis (of difficulties) The teaching function does not appear among their most favored activities Review of their vocational careers shows, however, that, in half of the cases, the teaching profession was chosen only after another vocation had first been tried. Again, one would expect Financial Speculation (activities aimed at financial profit) to appear among the most interesting functions of the business men This appears to be true, however, only in the initial *Inquiry*, and then only with the advertising managers. These facts suggest the desirability of an analysis of the items which were chosen to represent those functions to see to what extent they are truly representative.

An Item Analysis

In order to lay the groundwork for a possible revision of the *Inquiry*, the individual items have been examined, first, for the purpose of eliminating those that are non-differentiating, and, second, for the purpose of selecting those representative of the functions having the greatest predictive value. The non-differentiating items were located by the process of computing the proportional preference given to each item in the 1941 *Inquiry*. Of the 380 items in the *Inquiry*, 184 were preferred by 70 per cent or more of the subjects and were therefore eliminated Then, after the subjects had been grouped according to the vocation in which they were engaged in 1941, the relative popularity of the 196 items was computed for each vocational group. Examination of these proportions revealed 45 more items which were not distinctive of any vocation in which these particular men were engaged Their elimination reduced the total to 151 items One consequence of this procedure was the reduction of the number of items for some of the vocational categories to less than 16. The number of categories having at least 16 items was six, the largest number of items for any category being 37 The identifying names being used for these categories are 1) Commerce; 2) Merchandising, 3) Law and Social Science; 4) Medical and Social Work, 5) Physical Science, and 6) Literature. Because of the inequality in number of items under each category it became necessary to use proportions rather than summations of items in making comparisons of relative interest for the different activities.

The next step consisted in re-scoring the original responses on this selection of items to determine the extent to which the respondents' original answers would have predicted their final vocational choice At this stage the men engaged in Art and Music were eliminated because too few items were found which differentiated these vocations. This elimination reduced the test population to 112. (It should be noted that in order to re-score on these 151 selected items it was necessary to carry 104 pairs of items—including 57 undifferential items.) When the original inventories were re-scored on these 151 items it was found that 63, or 56.2 per cent of the men had the highest score in that category in which it would be most logical to classify the work in which they were actually engaged in 1941.

Further examination of the items suggested the possible advantage of making a still more drastic selection of items. In order to equalize numbers it was decided to select just the ten items in each vocational category which were most highly differential For this part of the study use was also made of the responses made on the original *Inquiry* by two additional groups of subjects, 27 students in the Hamma Divinity School, and 90 secretaries in city YMCA's in Ohio. The responses of these groups were studied with the possibility of finding items specific to the field of religious work. This quest proved, however, to be fruitless

With this final item selection, consisting of ten items for each of six categories, the Oberlin graduates who had responded to the second *Inquiry* were again tabulated on the basis of their first choice among the six vocational categories. (In addition to these 60 significant items it was necessary to retain 32 indifferent items, thus making a total of 92 items, or 46 pairs.) Classification of these men under these six categories on the basis of their actual final vocation shows that 56.1 per cent of them engaged in fields in the area in which they made the highest scores as seniors.

The vocational history of 47 other men who had filled out the *Inquiry* as seniors, but who had not responded to our second

Inquiry, being available, their original responses were scored with the 60-item form The results in the case of these men give a 66.7 predictive percentage, 32 of the 48 men having chosen an occupation in the one of the six areas for which they scored highest in the *Inquiry*.

As a final step in the validating analysis, the 60 items of the abbreviated *Inquiry* were classified on the basis of the twenty functions which had been used in constructing the original *Inquiry* When these were parcelled out in terms of the patterns characterizing the interests expressed by the various vocational groups (see Table 2), the number of items for each vocation proved to be too small to yield significant differential scores.

Summary and Conclusions

1. A validating follow-up has been made of the *Oberlin Vocational Interest Inquiry*, after an interval of fourteen years, based upon the responses of 68 per cent of the men tested in the year 1927–28. In an item analysis use was also made of responses from 90 secretaries of urban YMCA's and 27 students in a theological seminary.

2. When the correlation is computed between the rank order of interest expressed on the two occasions, 1927 and 1941, in the vocational functions sampled by the *Inquiry*, the coefficient for the entire group of 121 is 717

3. Examination of the patterns of preferred functions shows marked contrasts between the different vocational groups represented, and agreement with common sense expectations

4. Preliminary steps were taken toward a possible revision of the *Inquiry* in the form of item analyses aimed at eliminating non-differentiating items This process showed that, on the basis of the interest shown in the particular items used in this *Inquiry*, these men fell into six categories which have been labelled as follows· 1) Commerce, 2) Merchandising, 3) Law and Social Science, 4) Medical and Social Work, 5) Physical Science, and 6) Literature

5. The item study revealed 151 of the original 380 items which had been differentially selected by the members of these six vocational groups When the original 1927 *Inquiry* was re-scored, using only these 151 items, 56.2 per cent of the men made highest proportional scores in the vocational area in which they subsequently chose their life work.

6. The ten most highly differentiating items were selected for each of the six vocational categories, this constituting a further reduction of the total number of items to 60. Scores obtained from the original responses of the 121 men on this 60-item *Inquiry* give first choice to that category which most closely corresponds to the final vocational choice in the case of 56.1 per cent of them.

7 Finally, the responses made in 1927 by 47 men who did not answer the second request were scored on the 60-item *Inquiry*, and with this group it was found that, in two-thirds of the cases, the first choice was given that category represented by the vocation subsequently chosen.

8. It is to be regretted that the small size and specialized character of the population involved in this study precludes the use of more refined statistical procedures for item differentiation (3) The results, however, appear to harmonize with those obtained by assigning multiple regression weights to the items. They furnish further evidence, moreover, of the possibility of predicting vocational choices of college seniors within areas which stress particular functional interests.

REFERENCES

1 Fryer, Douglas *Measurement of Interests* New York: Holt, 1931, 38–40
2. Hartson, L D "The Vocational Stability of Oberlin Alumni" *The Personnel Journal*, VII (1928), 176–185.
3. Kuder, G Frederic. "A Review of Edward K. Strong's *Vocational Interests of Men and Women.*" *Psychometrika*, IX (1944), 147–148

GUIDANCE SURVEY OF STUDENT PROBLEMS

ELIZABETH GORDON ANDREWS
Florida State College for Women

THIS investigation was undertaken as a contribution to a curriculum study being made by a sub-committee of the Faculty Committee concerned with the functioning of personnel work and guidance on the campus of the Florida State College for Women. The students who secured the data were graduate students in a class in Personnel Work in 1941-'42.

These investigators visited as many of the girls as were found in their dormitory rooms when called upon. All students at the college reside in dormitories (or in approved student houses). The dormitories were divided between the two investigators. If a student was not in her room at the time the visitor called she was not included in the study. A total of 279 students, or 14 per cent of the total student body, were thus interviewed.

Each interviewee was presented with a sheet entitled "Guidance Survey." On this sheet were indicated the purpose of the study, a suggestive list of problems which the student might have had since entrance to college, a list of the possible sources of help that she might have consulted, and an example of the information desired. This study does not include the numerous times students are called to administrative and other offices. We limited our study to problems on which students voluntarily sought assistance.

Only those problems were recorded which assumed somewhat serious proportions. For example, the investigator would say, "Have you had any health problems since you have been in college? By health problems, we do not mean the times you may have gone to the Infirmary to have your throat swabbed, but instances in which your health was such as to definitely

affect your school work." Disciplinary problems included not only infraction of rules but also interpretation of rules. Extra-curricular problems included the case of need for dropping some activities as well as need for increasing them. All types of problems are defined on the sample sheet.

One of the chief factors in causing unreliability of report is due to the uncertainty of memory. For example, freshmen report homesickness as a problem more than upper-classmen. However, this may be partly due to the selection that occurs and the elimination of the homesick as well as the poorer students.

Sources of Help

Table 1 indicates the frequency with which the various offices or individuals on the campus served the students interviewed with regard to their problems.

TABLE 1

Frequency of Voluntary Requests for Aid

Number included in study, 279 Student body, 2007.

Source of help	Number of counsellors involved	Number of times sought*	Average per person
Academic Dean	4	105	26.2
Business Office	2	9	4.5
Faculty Advisers	60	99	1.5
Heads of Departments	25	30	1.2
Infirmary	2	77	38.5
Office of Dean of Students	2	10	5.0
Office of Director of Personnel	1	197	197.0
Registrar's Office	1	22	22.0
Religious Secretaries	4	45	11.2
Social Directors	10	85	8.5
Student Counsellors	50	62	1.2
Teachers (other than Faculty Advisers)	140	149	1.1
	301	890	

* Number of times sought refers to number of times aid was sought from the indicated sources, not to number of different students.

Both upper-classmen and freshmen report that they consult the Registrar's office solely regarding such problems as courses, credits, schedules, cuts and such matters as *pertain to* registration.

The Business Office is sought by both groups chiefly with

regard to financial matters related to college accounts. Twice they reported seeking aid in personal problems but in both these cases the student had a personal friend in this office.

The Dean of Student's Office functioned chiefly in three areas insofar as students voluntarily seek help from this office, extra-curricular problems, personal problems, and discipline.

The four areas in which the Personnel Office was most frequently sought were educational, financial, vocational, and placement problems. While personal problems were seldom listed this is perhaps due to the objective approach to problems in this office which avoids labelling personal problems as such. Placement is peculiarly a problem of upper-classmen, especially of seniors; hence it is to be expected that freshmen would not be concerned with this function of the Personnel Office. With this exception, upper-classmen and freshmen consulted the office most frequently on the same problem areas.

TABLE 2

Distribution between Freshmen and Upper Classes

	Freshmen		Upper-Classmen		Total	
	No.	%*	No.	%*	No.	%*
Academic Deans	25	7.5	80	14.4	105	11.8
Business Office	5	1.5	4	0.7	9	1.0
Faculty Advisers	57	17.1	42	7.5	99	11.1
Heads of Departments	4	1.2	26	4.7	30	3.4
Infirmary	35	10.5	42	7.5	77	8.7
Office of Dean of Students	3	0.9	7	1.3	10	1.1
Office of Director of Personnel	62	18.6	135	24.2	197	22.1
Registrar	9	2.7	13	2.3	22	2.5
Religious Secretaries	7	2.1	38	6.8	45	5.1
Social Directors	37	11.1	48	8.6	85	9.6
Student Counsellors	34	10.2	28	5.0	62	7.0
Teachers (other than Faculty Advisers)	55	16.5	94	16.9	149	16.7
	333	99.9	557	99.9	890	100.1

* The percentages given refer to the relation of the number of times aid was sought from the indicated sources by the designated group to the total times aid was sought by the group in question (Freshmen, Upper-Classmen, or Total).

The Infirmary was, of course, sought mostly in matters of health, with occasional related problems in which advice was sought from the physician or nurse.

The Office of the Director of Residence was sought a limited

number of times chiefly in regard to homesickness, roommate difficulty and occasional personal problems.

According to student reports, Social Directors are sought infrequently, but on a widely divergent list of problems including educational, religious, financial, health, vocational, personal, and disciplinary problems. It seems that the type of service the Directors in dormitories render most to the students lies in the field of personal problems.

Religious Secretaries are sought by students chiefly with regard to religious problems. The number of times the Secretary is sought seems to be much greater with upper-classmen than with freshmen. This may indicate that the Secretaries become increasingly helpful as they come to know the students longer, or it may indicate that religious problems are more prevalent among upper-classmen than among freshmen. A previous study on this campus would, however, indicate that the problems in this area are more prevalent in the earlier years.

The Student Counsellors were seldom sought regarding financial, religious, health, and placement problems. Their chief service includes assistance in educational, extra-curricular, and disciplinary problems.

Educational, personal, and vocational difficulties constitute the three outstanding problems about which Faculty Advisers are consulted. The small number of times Faculty Advisers were sought voluntarily was somewhat disturbing, but no doubt one cause of this is the difficulty of finding an hour at which the student and the Faculty Adviser have corresponding free time for conference. This again points to the need for Faculty Advisers to be given a reduced teaching load in order that they may devote more time to counselling.

It is evident from the student reports that Academic Deans assist students mainly with vocational and educational problems. This does not indicate vocational guidance so much as counsel with regard to pre-professional courses, such as pre-medical courses, courses for nurses, teachers, etc.

Freshmen consulted the Heads of Departments only on educational problems while upper-classmen conferred with them on problems of finance, extra-curricular activities, placement, and vocational guidance. This is probably due to the fact that freshmen have very little occasion to become acquainted with Heads of Departments until they have chosen their major field, which frequently is not until the latter part of the sophomore year.

Teachers other than Faculty Advisers make a valuable contribution to the guidance program by giving satisfactory help on educational and vocational problems. They are also often sought by upper-classmen regarding placement problems. They may be sought in any or all fields of guidance, but are not called on frequently in problem areas other than education and the vocations.

If the nine types of problems were ranked according to the frequency with which students reported that they sought help on these problems, the results would be as given in Table 3.

TABLE 3

Nature of Problems in Which Help is Sought

Freshmen			Upper-classmen		
Problem	No. times help sought	% of total*	Problem	No. times help sought	% of total*
1. Educational	139	40.3	1 Educational	199	34.8
2. Personal	60	17.4	2. Finance	70	12.2
3. Vocational	50	14.5	3 Vocational	70	12.2
4. Health	36	10.4	4 Personal	65	11.4
5. Finance	33	9.6	5 Health	43	7.5
6. Extra-curricular	11	3.2	6. Placement	43	7.5
7 Discipline	9	2.6	7 Religious	37	6.5
8. Religious	7	2.0	8. Extra-curricular	30	5.2
9. Placement	0	0.0	9 Discipline	15	2.6
	345	100.0		572	99.9

* The percentages given refer to the relation of the number of times help was sought in the indicated type of problems by the designated group to the total number of times help was sought on all problems by the group in question.

From Table 3 it is apparent that the problems that trouble both groups most are the educational ones, and this is the field in which students feel the most pressing need for assistance. Freshmen appear to need more help than upper-classmen with personal problems. Perhaps this is due to the more obvious sophistication of the upper-classmen. Vocational problems rank high with both groups, indicating that selection of vocation and preparation for a vocation are not confined to the first year but are also problems during the later years of college.

Problems of discipline, religious and extra-curricular matters are not so great a source of disturbance apparently as are other problems. Perhaps students fail to recognize them as problems and hence do not seek help with them. Students seek help most in educational, vocational, personal, financial, and health problems. The attention of Counsellors might be directed toward improvement of assistance in those areas in which students report seeking help most frequently, and further investigation should be made to determine whether problems actually exist in other areas where students are unaware that they need counsel.

TABLE 4

Help Voluntarily Sought by Students (according to records) in Personnel Office

	No. of requests	Record percentage	Survey percentage
Financial	426	35	41
Educational	362	30	27
Personal	131	10	02
Vocational	109	08	15
Placement	109	08	14
Health	65	05	01
Extra-curricular	34	03	00
Religious	7	01	00
	1243	100	100

During the interviews, forty-eight students reported that they had had problems about which they had not sought help. Twenty of these cases were personal problems, the remainder being distributed over all the other fields. It is apparent that personal problems constitute the outstanding area in which students realize that they need guidance but hesitate to ask for help.

Out of the sixteen possible sources of official assistance available to students on the campus, this survey reveals that the six offices or individuals sought most frequently by students include in order of decreasing frequency (1) Office of the Director of Personnel, (2) Teachers other than Faculty Advisers, (3) Academic Deans, (4) Faculty Advisers, (5) Social Directors, and (6) the Infirmary.

In Table 4 it will be noted that the actual problems in which help was sought by students during the period of time covered by the survey are very much in the same order as reported by the students, indicating that the reports were fairly accurate and that the sampling by the method used was quite representative. The only change in order would be placing the Personal problems below Vocational and Placement problems. As stated previously, this is probably due to the approach in the Personnel Office which seeks to minimize the "personal" problem in favor of a more objective approach. At the same time the confidential records of the office indicate the personal problem involved.

INTERPRETATION OF THE KUDER PREFERENCE RECORD FOR COLLEGE STUDENTS OF HOME ECONOMICS

RUTH T. LEHMAN
The Ohio State University

FOR the past several years, the *Kuder Preference Record* has been used experimentally with freshman students, in connection with a rather extensive guidance program in the School of Home Economics of the Ohio State University. This experimentation has shown that the *Record* makes a helpful contribution to the total picture of the individual which is needed by the counsellor, suggesting both vocational and avocational interests.

However, interpretation of interest profiles has presented serious difficulties. Some have assumed that, since the home economics curriculum has traditionally been built upon a base of natural science, performance interests in that area should be strong; that, in fact, students who have little interest in scientific activities should be barred from entering or continuing in the field of home economics. Yet interest scores in science and marks received in science courses by home economics students have been found frequently to be in direct contradiction. Similarly, one might expect to find a pattern of interests which is characteristic of home economists. Such has not proved to be the case. The weakness of the assumption is evident when one surveys the diverse fields which the graduate in home economics may enter. She may become a teacher, an extension worker, a hospital dietitian, a foods manager in a restaurant, or she may enter one of the numerous divisions of business or of social service. Obviously, it would be dangerous to base interpretation of an interest profile either on one's prejudices, or on a guess as to areas which are significant for specific occupations requiring a home economics background.

Such difficulties in interpreting the *Record* blocked its effective use with college freshmen in home economics. Therefore, the test was given experimentally to a sample of home economics teachers, dietitians, and home economists who were employed by utility and household equipment companies Findings from this preliminary work indicated that there were real differences in interests of the three groups and encouraged an extension of the study. Other cases were added during the school year of 1943–44

The Study Reported

To date the *Record* has been checked by 239 home economists. These are distributed as follows. 125 secondary-school teachers, 42 hospital dietitians and 72 home economics women in business Of the teachers, 51 were volunteers from a summer institute held for Ohio vocational teachers; 74 others in the State were reached through meetings called by their city supervisors The hospital dietitians were secured through group meetings in several Ohio cities, and through a random sampling of the Ohio Dietetic Association. The business women were from various parts of the United States. They represent a random sampling of the extensive membership list of the Home Economics in Business department of the American Home Economics Association, supplemented by a smaller number of Ohio women not included in this list. Of the business women, 10 were engaged in restaurant or tea-room management, 11 in journalism, and 20 in foods promotion or experimental work, and 31 were employed by utility and household equipment companies All three groups of home economists represented a wide range—but a fairly good balance—in terms of experience in their respective fields (Table 1).

Form B of the *Kuder Preference Record* was used, except in the early part of the study when only Form A was available In order that all data might be used, the raw scores secured on Form A were changed to comparable scores on the later form by means of a transmutation table provided by the author of the test. Medians were found in each interest area and their percentile equivalents located on a profile sheet. Mention should be made also of the fact that only a part of the total group received scores in the mechanical and clerical areas, since these areas were not included in the early form (See footnotes to the various tables)

TABLE 1

Distribution of 239 Home Economists According to Years of Experience in Their Respective Fields

Years of experience	Teachers		Hospital dietitians		Business women	
	No.	%	No.	%	No.	%
1–4	27	22	15	36	23	32
5–9	17	14	12	29	18	25
10–19	37	30	8	19	22	31
20–	43	34	2	5	7	10
No data	1	*	5	12	2	3
Total	125	100	42	101	72	101

* Less than 1 per cent.

While the interest patterns of different persons as revealed by the test were highly individual, there appeared to be quite distinct differences among the three groups of home economists included in the study (Table 2). The teachers and dietitians had high social service interests; the business women were somewhat lower. The dietitians were outstanding in science interests; the other two groups were lower, although still above the average for college freshmen The dietitians were also much higher in the computational area than were the teachers and business women. The teachers surpassed the others in artistic interests, the business women, in the persuasive or selling area

TABLE 2

Median Percentile Scores of Home Economists in the Nine Interest Areas of the Kuder Preference Record

Professional group	Mechanical*	Computational	Scientific	Persuasive	Artistic	Literary	Musical	Social Service	Clerical*
Teachers ... (N=125)	57	43	63	12	71	37	29	83	34
Hospital dietitians (N=42)	45	63	82	7	52	43	39	80	34
Business women ... (N=72)	49	43	60	51	59	53	29	62	23

* In the mechanical and clerical areas, scores were not available for 51 of the teachers, 12 of the dietitians and 17 of the business women.

In other words, the areas of highest interest for the teachers were artistic and social service, median percentile scores in those areas being 71 and 83, respectively In fact, over three-fourths of the teachers (77 and 78 per cent, respectively) had scores at or above the median for college freshmen in these areas. There was some evidence also of scientific interest, but this was not nearly so high as in the case of the dietitians A breakdown of the teacher group in terms of years of experience shows that this pattern of activity preferences is consistent (Table 3). An apparent gain in artistic interests and loss in the social service area in the case of the teachers having the longest term of service may be revealing.

TABLE 3

Median Percentile Scores of Home Economics Teachers when Classified as to Years of Experience

Years of experience	Mechanical*	Computational	Scientific	Persuasive	Artistic	Literary	Musical	Social Service	Clerical*
1–4 (N=27)	73	41	64	12	62	37	35	89	39
5–9 (N=17)	67	31	69	3	68	20	34	83	13
10–19 (N=37)	63	46	51	15	67	37	22	85	39
20– (N=43)	55	50	64	13	75	43	25	78	37

* Scores in mechanical and clerical areas were not available for 51 of the teachers. Therefore, in the 1–4 yr group, N=4; 5–9 yrs. 6; 10–19 yrs. 22, and 20 years and over, 41.

The hospital dietitians, though fewer in number than the teachers, also presented a clear pattern of interests Outstanding areas were computational, scientific, and social service Median percentile scores in these areas were 63, 82, and 80, respectively. Moreover, around three-fourths of the dietitians had scores above the 49th percentile (69, 76, and 79 per cent in the respective areas).

Interestingly enough, though the business women differed in certain respects from the teachers and dietitians, they showed no definite preference pattern Interests in as many as five or six areas appeared to be similar (Table 2). However, when this group was broken down in terms of occupation within the business field, differences did appear (Table 4) Certainty as to how real these differences are will depend upon evidence provided by additional cases, for the numbers representing the different occupations were small. Interests apparent in restaurant and tea-room managers were computational, scientific, and artistic (71, 64, and 76 being the respective medians, and 70, 90, and 60 being the respective per cents above the 49th percentile norm). Similarly, apparent interests of the women in foods promotion and experimentation were in the scientific area (median percentile 78, with 80 per cent above the 49th percentile norm), of women in home service and equipment, in persuasive (selling), social service, and probably mechanical areas (medians of 68, 68, and 62, with 68, 68, and 79 per cent of the cases, respectively, above 49). Interests of the journalists were

TABLE 4

Median Percentile Scores of Business Home Economists when Classified as to Business Group

Business group	Mechanical	Computational	Scientific	Persuasive	Artistic	Literary	Musical	Social Service	Clerical
Restaurant and Tea room ... (N=10)	57	71	64	47	76	32	18	50	32
Foods Promotion (N=20)	45	59	78	20	48	58	34	64	23
Home Service and Equipment .. (N=31)	62*	31	54	68	58	49	25	68	20*
Journalism (N=11)	36	38	47	48	78	71	38	57	26

* In the mechanical and clerical areas, scores were not available for 17 of the home service and equipment women.

centered in literary and artistic areas (medians of 71 and 78, in each case 82 per cent being above 49) *

Use of the Findings in the Guidance of College Home Economics Students

The foregoing findings suggest certain conclusions concerning the interpretation of interest profiles for home economics students

1 The study does not reveal a pattern of interests which is characteristic of home economists as a group

2 However, the student who is considering home economics as a field or who is already enrolled in it can profitably compare her own interests with the outstanding preferences of those actively engaged in various professions requiring a home economics background

3 Such comparison should serve as the basis for a discussion of specific activities in the field being considered which call for certain interest patterns For example, the scientific interests of the hospital dietitian are clearly used in the planning of special diets based on exact scientific facts, her computational interests in the keeping of necessary accounts, the estimation of diets of given composition, and the ordering of food supplies. Similarly, the artistic and social service interests of the home economics teacher should be related to creative activities in home making and to a desire to work with others and help them.

4. Ideally such discussion should eventually lead to exploratory experiences on the part of the student to discover her abilities and whether she really enjoys such activities Perhaps this should follow particularly in those cases where test results and professional ambitions are in evident conflict.

5. In any case, however, the interest profile should be used in relation to other factors,— the student's personality, intelligence, special aptitudes, handicaps, health, and educational and experiential background. Left to herself, the student is likely

to rely too much on test scores Often she would like a test which would make decision on her part unnecessary. On the other hand, the inexperienced counselor with too great a respect for tests may use the profile in a dogmatic fashion or without supporting evidence.

6. Furthermore, it is important for both student and adviser to interpret the *Preference Record* in terms of what it claims to measure, namely, activity interests, not ability, nor appreciation. This has been found to be doubly important to remember in the artistic, writing, and musical performance areas, where ability or appreciation are so readily assumed to be the interpretation.

7. The counselor must keep in mind also that, although certain interest patterns appear in studies of given occupational groups, profiles are after all highly individual, reflecting both possible vocational and avocational interests, and may change with experience and maturity

Further Work Needed

This account of the present investigation is in the nature of a progress report. Need for further study is evident. In the first place, the validity of the tentative findings for dietitians and home economists in business must be tested by increasing the number of cases Then, too, data from other related professions—such as home economics extension, and the textiles and clothing area of business—should be added, if the findings are to be of greatest use to the counselor of home economics students

Finally, certain general questions need to be studied. Among them are two which face every counselor who attempts to interpret students' interest scores One of these is the big question of how much student activity preferences change during the four years in college and what are important factors in such change The other is the basic question as to whether the pattern of interests revealed in a study of professional workers is closely related to experience in and knowledge of the job, or is more a reflection of the selective effects of the specific undergraduate curriculum which was followed.

---

* Throughout the study, percentile norms for Form B as established for first year college men and women were used, rather than the more recent norms for high-school girls. The latter table would show the home economists higher in general than they are recorded in this paper in mechanical and literary interests, and somewhat lower in social service interests. It would also give even higher medians in some areas for various sub-groups.

THE USE OF STANDARDIZED READING TESTS IN TEACHERS COLLEGES[1]

ALFRED M. EWING
Texas Wesleyan College

WHAT is the most efficient way to use the results of reading tests? What is the most usable test in the reading field? What disposition is made of reading test scores? In an attempt to answer these and similar questions concerning the use made of standardized reading tests in teachers colleges, data were gathered by means of a mimeographed two-page twenty-seven-item questionnaire which was sent to 167 teachers colleges One hundred three returned the questionnaire, making the percentage of returns, 61.7%, more than twice that of similar surveys in colleges and universities No reading tests were given in 17 of these institutions, or a program was just beginning. Thus the data here given are from 86 institutions answering questions concerning the use made of reading tests and the use and interpretation of the results.

The 86 teachers colleges using reading tests reported the use of 27 tests. Four tests, the *Iowa Silent Reading Test,* the *Cooperative English Test (Form C2, Reading Comprehension),* the *Nelson-Denny Reading Test,* and the *Entrance and Classification Test for Teachers Colleges Entrants. English Test,* are used much more extensively than are any of the other 23 tests. In fact the four tests make up nearly three-fourths of all the tests, their respective percentages being, 35.5%, 20.0%, 10.3%, and 6.8% Thus limited use of a wide variety of tests is accounted for by local use of the test, use by its author, or some other specific reason and not by general considerations such as ease of scoring, popularity of the test, or number of skills or abilities tested.

Reasons for Selecting Tests

The ability to diagnose reading difficulties is given as the main reason by 16% of the teachers colleges for selecting the reading tests which they administer to their students. Following closely is the fact that it "meets their needs," 14%. Simplicity of scoring, 10%; high reliability and validity, 9%, and recommendation or selection by authority, 9%, are other reasons frequently given by teachers colleges for selecting specific reading tests. Twenty-two other reasons for selecting reading tests are mentioned once or twice.

Practices in the Techniques of Testing

Four departments, psychology, testing, English, and personnel and guidance, administer the reading test in half of the teachers colleges. In the other half reading tests are administered by ten other agencies.

The faculty ranks first as the group most often *scoring* reading tests in teachers colleges. This includes such departments as English, psychology, etc. There seems to be, from the data obtained, no one department by which the scoring is generally done outside of the faculty and assistants.

The personnel and guidance office ranks first as a place where reading test scores are *filed* The deans' and registrars' offices rank next while the English department ranks fourth.

The personnel and guidance office also ranks first as a place where the scores are available or *given out.* The adviser and dean rank second and third, respectively. In eight schools scores are not given out, while in many other schools more than one department gives out the test scores

Fifty of the 86 schools give test results to the students. In three of the 50 schools the test scores are given only occasionally to the student and in six other schools they are available to him

The Use Made of Reading Test Scores

The purpose of obtaining reading test scores, as stated by 40% of the teachers colleges, is to find the students who need help in reading. Twenty-two per cent more say for diagnosis, while 16% give appraisal of the student's reading ability as the purpose. An analysis of these purposes shows a definite need for a good achievement test in reading as well as a diagnostic test.

---

[1] This article is based on a study done in partial fulfillment of the requirement for the degree of Doctor of Education at Colorado State College of Education, Greeley, Colorado.

The most common reason for giving reading tests is to select those students needing remedial help in reading. But an examination of the purposes given for obtaining reading test scores reveals other reasons for giving reading tests, such as determining daily lesson assignments, improving study habits and the like.

Half of the teachers colleges supplying data for this study use testing *alone* for selecting those students who need help in the improvement of reading A fourth more use testing plus other items, while another fourth use no tests. This means that standardized tests were used in 75% of the schools for selecting students needing remedial reading.

Just how the results of reading tests are actually used may be best described by including excerpts from reports submitted by a number of teachers colleges.

In College A all students who are doing unsatisfactory work are interviewed by the personnel director The reading test percentile ratings are compared with the percentile ratings on the psychological examination. If there is a marked discrepancy the student is told that he will be given help in the improvement of reading if he wishes it. Help is not confined, however, to those having such a discrepancy.

College B uses test scores to discover personal deficiencies in reading ability, to assist in placement of students in English classes according to reading ability level, to assist in counselling students; to assist students to help themselves; to assist in evaluating curriculum needs; to assist the students in detecting their own strengths and weaknesses

At College C the scores are incorporated in the "Guidance Tests Record Profile" of the Cumulative Personnel Record. They are given to the students for the purpose of self-appraisal in terms of their own group and the national group. Open and frank discussion is entered into in regard to the possible reasons for a lower score in one test if there is a difference. Occasionally students reveal special conditions which are responsible for low

scores. Very few bona fide low scores exist and they are usually caused by lack of speed.

In College D test scores are used as one of the bases for advising students relative to the academic load they should carry. Instructors may consult records to determine probable student abilities. Deans and Advisers find the records valuable in giving them an index of the students' probable success, and a limited number of superior students are granted a partial tuition exemption. They prefer to select superior students rather than attempt to save (for teaching) those of inferior ability and attainments

As an example of the use of reading tests in solving personnel problems, one of the teachers colleges makes the following statement from a five-year survey in the college. The college records data from several tests including one on the measurement of ability in reading. From the great variation in reading ability discovered, certain questions arose. What curricula do these poor readers enter? How long do they remain in college? What kinds of grades do they make?

The findings were: *Curricula*—physical education, agriculture, home economics. No one chose art, biology, chemistry, English, music, or physics as his major field. *Years in college*—the mortality rate of poor readers is exceedingly great. Only 70% are in school at the end of the first year, 45% at the end of the second year, 20% at the end of the third year, and 15% at the end of the fourth year. *Grades*—those that remain make grades comparable to grades of other college students.

Plans Used in the Remedial Reading Program

Twenty of the 86 teachers colleges have specialized reading classes for improving the reading ability of the students; fifteen have both a specialized reading class and a reading clinic, while eleven report only reading clinics Thirteen schools use special sections of certain courses, while nearly equally as many apparently give no concern to reading. The clinic and special classes, however, account for over 51% of the 86 schools whose attention is drawn to actual improvement.

Small schools, that is those below 500 enrollment, give slightly more attention to reading than do larger schools, but the difference is not significant

How some of the plans for improving reading actually operate is seen from the examples given, one each for a reading clinic, a laboratory, an individual conference method, and an orientation course.

From a *reading clinic*—One of the least gratifying aspects of the program has been the uncertainty of results. In some instances visiting classes all day, conducting a conference with teachers after school, and spending extra hours in conversation have produced no tangible evidence of change in the instructional program. However, most teachers show a keen interest in the questions discussed.

But tangible results are sometimes, oftentimes, almost overwhelming. Witness the case of Miss F, a college girl who made the following record from the analysis of her eye movements as photographed by the ophthalmograph.

1944	*Jan. 7*	*Feb. 26*	*May 16*
Rate of reading (wpm)	226	286	320
Fixations	112	100	84
Regressions	30	20	16
Average word span recognition	.89	1.6	1.2
Average duration of fixation in seconds	.24	.21	.2

The increase of 94 words per minute can be readily appreciated when practically applied to the work of a student. That means 5640 more words per hour provided the rate can be maintained, and assuming she reads five hours a day, five days a week, her *increased* reading achievement rises to an impressive 141,000 words per week, or over four books of 30,000 words each.

A *laboratory reports* that the word "laboratory" was used because it implied a broader, more general use than the word "clinic" and did not limit the function to diagnosis and correction. The members of the class learn how to set up and maintain case histories, how to operate certain simple equipment, how to administer, score, and interpret many kinds of tests; and how to locate and assemble this essential information and data in practical situations. Students learn to think of tests as aids and guides, not as infallible conclusions.

In a case of *individual help* the director of reading improvement listed the lowest quarter of the freshman class for special help as indicated by results of the *Iowa Reading Test*, then interviewed the staff for freshman reading assignments in college subjects, read the assignment himself, and made objective comprehension questions on the content. The students read the assignments silently in the presence of the director, they were told to read carefully but were urged not to hurry. Then they answered from what they remembered At the end of a time another form of the reading test was given. The improvement was statistically significant. This type of work is good but it is not always possible to schedule a special class. Little extra time was required for they read the assignments that they had to read anyway

One teachers college used the *orientation* course to improve reading ability under the following topical ideas: Discussion of reading as a tool, relationship of speed and comprehension; administration of timed informal tests—the students were required to find specific answers to questions in their assignment; use of association tests, in which the student is given a word or two and writes all he knows about these words during a minute or two, development by students of informal tests, true and false, completion, etc , use of assignments to provide for individual differences by noting books that are easy or difficult; check of visual and auditory efficiency of students whose rate is low.

Recommendations

On the basis of the data obtained from the cooperating teachers colleges, the following recommendations are made·

1. More teachers colleges should stress remedial reading. Forty-one of the 86 schools said they felt their reading program was fairly satisfactory, 33 were dissatisfied, eight were undecided, three did not indicate, and one school said it had no problem in reading.

2. There is a definite need for both a good achievement test in reading and a good diagnostic test.

3. Remedial reading should have a classroom approach as well as a clinical approach for both speed and comprehension.

4. Marked improvement can be made without the use of expensive apparatus such as metronoscopes, tachistoscopes, films, and corrective reading manuals, by giving a little time to practice, drawing attention to a few efficiency devices in better reading, and using mimeographed materials.

5. Follow-up tests should be administered to determine the permanency of the results of remedial reading programs.

6. With less attention than one realizes reading can be improved. For example, the student's attention can be drawn to discarding useless motions, such as pointing, head or lip movements, or whispering, to making fewer and shorter fixations, to seeing phrases and sentences at a glance, to anticipating what the author is going to say instead of passively attempting to absorb his ideas, and to practicing reading at a "faster than convenient" rate.

MEASUREMENT IN THE CONTINUOUS SELECTION AND COUNSELING OF STUDENTS IN A COLLEGE OF PHYSICAL EDUCATION AND SOCIAL WORK

J. E. TODD*
The Institute of Paper Chemistry, Appleton, Wisconsin

SPRINGFIELD COLLEGE is seeking to solve through research and measurement some of the problems confronting it in the selection and training of leaders of youth for positions as teachers, program directors, and directors of social agencies, such as the Y.M.C.A., Boys' Clubs, settlements, Boy Scouts, camps, parks, recreation centers, and other agencies of informal education. Those in charge of the personnel services recognize two clear lines of responsibility. One is the function of selecting and making sure that only those who are competent are admitted and certified to the professional ranks. The other is the function of counseling the individual in order that he will find the right road for himself and make progress toward his professional goal.

1 *Discovering Candidates* (14)

Two-thirds of the suitable candidates for admission are discovered by present students and alumni of the college who as physical directors, Y.M.C.A. secretaries, high-school teachers, and social workers are in close association with high-school boys. Rather complete admission data are required—viz., high-school transcript, ratings by principal and others, scholastic aptitude test scores, medical history, and personal history. Each candidate is interviewed by at least one member of the Alumni Admissions Committee or the College Admissions Committee, persons who can judge the applicant as a potential professional man. Admission is based on intellectual competence, personal and social adjustment, leadership ability, health, physical qualifications, finances, and interest in a profession for which the college educates.

2. *Continuous Counseling and Selection*

The college provides for continuous counseling services through the college course beginning with the orientation program of "Freshman Days" (4), which includes getting settled in the dormitory, registration, testing, social affairs, and explanations of procedures and traditions, and which is carried out to a large extent by a committee of upper-classmen.

The Common Freshman Year (13) was begun in 1937 to provide general education to meet the needs of students in adjusting themselves to the college community and life's responsibilities in a career of youth leadership.

A series of psychological and educational tests of abilities, achievements, adjustments, and interests (1) and a series of physical tests of organic and physical efficiency, motor ability, posture, and nutrition (7) form the basis of a thorough program of selection and counseling throughout the student's career in college.

In the group guidance course, "Introduction to Education," carrying six semester hours credit for the year, individual results of the tests are given to the students and interpreted as an aid to self-appraisal and self-direction. Along with these are considered the student's adjustment to college life, the meaning and purpose of higher education, methods of study, efficient reading, and vocational and educational planning. In the winter term, consideration is given to health adjustments and personal hygiene; and in the spring term, to religious, ethical and socio-vocational adjustments. Against all these, the student checks his own qualifications and maps out his curricular plans for the next year and the remainder of his college course. He makes application for admission to the professional division of his choice and all admissions and freshman-year records of the student, including a profile of all aptitude, achievement, personality, and physical tests, accompany the application to the professional division with the recommendation of the student's advisor and the Chairman of the Freshman year. Each freshman is interviewed by the Divisional Admissions Committee and is accepted or counseled as to dropping out of college or transferring to another more appropriate college. Thus the orientation course, the testing program, and counseling not only form the bases of sound guidance to the student, but also give many opportunities for objective appraisal of the student and careful selection for the professional divisions.

Students enter the professional divisions at the beginning of the Sophomore year. Each man is assigned to a special counselor who helps him plan his program for the next three years and supervises his work during the year. All men carry a full program of Liberal Arts and Science courses selected to serve preprofessional values. Continued development in recreational skills is required. At the end of the year, students participate in the College Sophomore Testing Program. The Sophomore year is a transition from general to professional education. Advisors and divisional officers using grades, test scores, and ratings, appraise carefully each man and make decisions regarding him. The promising men are encouraged to continue, and the doubtful men are advised to plan other channels of endeavor.

The individual counseling by faculty members of the professional divisions continues in the Junior year. The course of study includes several professional courses, and two or three advanced courses in liberal subjects are required. Supervised field work begins in the local schools and social agencies, which gives practical experience and tests the professional aptitude and interests of the student. In the Senior year, counseling by divisional advisors, usually by the professional divisional directors themselves, continues. The proportion of professional study increases and the program becomes almost entirely professional. Students can concentrate in their particular fields of interest. Field work increases and emphasizes experience in supervision and is more critically appraised.

Placement plans by the four-year men are made and credentials are filed in the college placement office. Correspondence

---

* Formerly Director of Admissions and Student Relations, Chairman of the Freshman Year, Springfield College, 1937–1941.

and interviews with employers are arranged The recommendation of men for particular jobs is the responsibility of the professional staff of each division. Final decisions and plans regarding the fifth-year program or other graduate study are made by those students who are qualified

Those continuing for the fifth year increase their professional study and field work and complete their theses or research projects if they are candidates for the Master's Degree. The same general program continues on this level with even closer faculty-student relations in all professional and educational matters.

Every student is expected to spend one or more summers in work experiences of educational significance. Most of the students become camp counselors, camp directors, playground and beach supervisors, or serve on the regular staff of some social agency which does not have a special summer program Reports of these summer's experiences are required

The intimate social relations and physical conditions of dormitory life are influential factors in the development of students Supervision of the dormitories is in the hands of a committee composed of faculty and students Upper-class counselors, chosen by the dormitory committee, counsel students concerning dormitory adjustments and personal questions. Meetings of these counselors are held regularly, and a program of educating the counselors to more effective service is carried on.

Thus throughout the college program, continuous selection and counselling procedures are in effect and make use of psychological and physical measurements.

3. *Descriptive and Predictive Values of Tests and Measurements*

Measurement is useful in making decisions in selection and counseling because of its descriptive and predictive values The following pre-war data have been chosen as being more typical of the normal functioning of the college than data available in the war years. The data in Table 1 give a picture of the *average* freshman student. These averages are the result of

the various selective factors leading to enrollment in Springfield College These selective factors appear to be operating successfully In the opinion of the administration, the averages obtained represent the type most suitable in terms of the function of the College.

TABLE 1

Test Score Description of the Average Springfield College Freshman Student

A Pre-admission Data (14), class entering in 1940

Freshmen admitted	
[illegible]	132
[illegible]	52
[illegible] in freshman class	
[illegible]	60
[illegible]	80
[illegible]	88
[illegible]	10
[illegible]	2
Percentile (national norms) for Liberal Arts College Freshmen	169 72 52

Principal's Ratings (9)[1]

	Mean	S.D
Integrity	6.79	.70
Social [illegible]	5.75	.98
[illegible]	5.57	1.05
[illegible]	5.28	1.14
[illegible]	4.64	1.12
[illegible]	4.45	1.37
[illegible]	4.24	1.00

B. Post-admission Data (1)

1 Psychological and Educational Tests—class entering 1940

	N	*Mean*	*S.D*	*Percentile (National College Freshman Norms)*
A C.E Psychological Examination (1938 edition)				
Linguistic	131	49.14	14.05	50
Quantitative	131	28.32	6.99	65
Total score	131	77.00	18.20	58
Cooperative Literary Comprehension Test, Form C				
Speed	130	58.90	7.48	60
Level	130	56.90	6.60	50
Cooperative English Test (*O.M*)				
Total	130	50.40	8.90	30
Cooperative General Culture Test				
Social Studies	131	36.20	18.50	52
Foreign Literature	131	31.65	13.30	40
Fine Arts	131	30.55	16.00	37
Science	131	49.10	14.10	60
Mathematics	131	19.52	8.00	60
Total Score	131	156.85	55.50	48

---

[1] The scales on which the ratings are based extend from 0 at the lower end to 8 at the high end.

TABLE 1 (*Continued*)

Allport-Vernon Study of Values				
Religious	131	34.66	7.08	85
Political	131	33.61	6.51	75
Social	131	32.10	5.34	70
Theoretical	131	29.19	5.94	55
Economic	131	29.08	5.58	55
Aesthetic	131	21.34	6.33	20
Bell Adjustment Inventory				
Home	114	3.94	3.40	"good"
Health	114	5.29	3.14	"average"
Social	114	9.26	6.40	"aggressive"
Emotional	114	6.63	5.22	"average"

Strong Vocational Interest Blank (Revised Form for Men)

	Average rating of three classes entering in 1938, 1939, 1940
Artist	C+
Bacteriologist	C+
Architect	C+
Physician	B−
Dentist	C+
Mathematician	C+
Engineer	C+
Chemist	C+
Production Manager	B−
Mathematics—Science Teacher	B
Y.M.C.A. Physical Director	B+
Personnel Director	B−
Y.M.C.A. Secretary	B−
Social Science Teacher	B
City School Superintendent	C+
Minister	C+
Musician	[illegible]
C.P.A.	C+
Accountant	C+
Office Worker	B−
Purchasing Agent	C+
Sales Manager	B−
Real Estate Salesman	B
Life Insurance Salesman	B−
Advertising man	B−
Lawyer	B−
Author—Journalist	C+

2. Comparison of Certain Achievement Test Results

	National Norms for Liberal Arts Colleges, Median Score, Percentile
Cooperative English Test, Form O.M (1)	
Class of 1942	
September, 1938, Entering Freshman	35
March, 1939, Freshman	53
April, 1940, Sophomore	53
Class of 1943	
September, 1939, Freshman	31
April, 1941, Sophomore	61
Cooperative General Culture Test	
Class of 1942	
September, 1938, Freshman	55
April, 1940, Sophomore	60

TABLE 1 (*Continued*)

Class of 1943	
September, 1939, Freshman	50
April, 1941, Sophomore	68

3 Physical Tests (7) for classes entering as indicated

Test	*Mean*	*S.D.*
Height—		
1939–40	69.10 inches	2.47
1940–41	68.71 inches	2.36
Weight—		
1939–40	156.84 pounds	17.00
1940–41	158.82 pounds	16.00
McCurdy-Larson Organic Efficiency Test (circulatory respiratory function)		
1938–39	185.20	89.08
1939–40	202.05	88.50
Cozen's General Motor Ability Test		
1939–40	345.06	80.36
1940–41	339.92	71.49
Springfield General Motor Capacity Test		
1939–40	236.53	15.54
1940–41	242.39	18.10
Springfield Muscular Strength Test		
1939–40	293.63	59.70
1940–41	299.71	89.18

Tests of agility, large and small muscle coordination, posture, and nutrition are also administered during the first term. Analysis of data indicates that the freshmen who expect to major in health and physical education are superior to those majoring in the other professional divisions in general motor ability and capacity, in general strength development, and in motor development.

A factor analysis of strength tests resulted in two components which can be measured by a few simple tests, viz., dynamic strength, which describes a composite motor ability criterion to the extent of 68%, and static dynamometrical strength, which was responsible for 24% of the reduction (8).

A factor analysis of 18 cardio-vascular test variables and the *McCurdy-Larson Test* and the *Tuttle Test* isolated eight principal components, viz., three concerned with pulse rates, two with blood pressures, two with pressures referring to postural changes, and one with circulatory recovery to exercise (11).

Other factor analyses of variables in specific sports, such as soccer, basketball, and other skills have been made by Kostantinov (6), Tibbetts (12), Hunsicker (5), and others. Important theoretical and practical studies of posture and aquatics have been made by Cureton and of endurance and nutrition by Karpovich.

Table 2 gives some idea of the predictive relations between test scores and achievement. Measurement has been useful in predicting individual and group achievement

TABLE 2

Prediction of Fall-Term Freshman Grade Index and English Grades from Pre-Admission and Post-Admission Data

1. Fall-Term Grade Index (9)	
Correlations between	
0, Fall-Term Freshman Grade Index, and	
1. 1937 *A.C.E. Psychological Examination* (pre-admission)	.393
2. Percentile rank in high-school graduating class	.350
3. Rating on *A.C.E. Personality Report, Revision B,* "Program and [illegible]	[illegible]
4. [illegible]	[illegible]
5 [illegible]	[illegible]
6. [illegible]	[illegible]
Multiple Correlations	
R[illegible] = 691	
R[illegible] = 675	
R[illegible] =.731	
0, Fall-Term Freshman Grade Index (9), and	
1. Combined Fall and Winter Terms (Freshman Year) Grades	.925
2. End of First-Year grades	.924
3 End of Second-Year grades	772
4. End of Third-Year grades	.725
5. End of Fourth-Year grades	752
Cooperative English Test (10)	
Correlations between	
0, Cooperative English Test, Total Score, and:	
1 [illegible]	.327
[illegible]	199
[illegible]	
[illegible]	.485
[illegible]	
[illegible]	191
[illegible]	630
[illegible]	488
[illegible]	.294
Multiple Correlations	
R[illegible] = 706	
R[illegible] = 715	
R[illegible] =.721	
R[illegible] =.714	
R[illegible] =.692	
R[illegible] =.691	

TABLE 2 (*Continued*)

3 Fall-Term English Grades (10)	
Correlations between	
0, Fall-Term Freshman English Grades, and:	
1 High School English Grades	497
2. Percentile Rank in High-School Class	157
3 Size of High-School Class	150
4 Completion part-score 1937 *A.C.E. Test*	.539
5 Artificial Language part-score *A.C.E. Test*	.179
6 Opposites part-score *A.C.E. Test*	.511
7 Total Score 1937 *A.C.E. Test*	.472
8 Other Language in High School	.257
Multiple Correlations	
R[illegible] = 855	
R[illegible] =.855	
R[illegible] =.723	
R[illegible] = 708	
R[illegible] =.619	
R[illegible] =.618	

4. *Effectiveness of Program*

The ultimate test of the validity of the whole selection, counseling, and academic program is its effectiveness in preparing men for youth leadership Pertinent to this subject are the data on academic mortality, placement, and professional stability given in Table 3. Data on Springfield College students for Part 1 of this table covered 8 years. The initial placement data in Part 2 of the table covered approximately 900 placements over 20 years The data on permanence of the field of initial placement were collected 5 to 20 years after the initial placements involved. Strictly speaking, of course, a rigorous test of the effect of the program would require conducting a controlled experiment of such magnitude as to be impractical. However, it seems reasonable to believe that the relatively low academic mortality and the high placement and success in social work may be attributed in part to the use of measurement and prediction procedures in selection and counseling

TABLE 3

Summary of Academic Mortality, Placement, and Professional Stability

1 Academic Mortality of the Classes of 1934, 1937, 1940, 1941, and 1942 (2)	
Loss during and at end of Freshman Year	
[illegible]	24%
[illegible]	35.6%
[illegible]	31.5%
[illegible]	
[illegible]	21%

Of Freshmen lost from Springfield during first year, classes of 1941 and 1942	37%
[illegible]	30%
[illegible]	23%
[illegible]	
[illegible]	
[illegible]	
Springfield	55%
Public Controlled Institutions (16)	27 7%
Privately Controlled Institutions (16)	34%
2. Placement (3)	
Initial employment of 921 graduates during years 1921 to 1940	
By position	
1 [illegible]	65.2%
2. [illegible]	23 1%
3 [illegible]	2.0%
4. [illegible]	6.0%
5. [illegible]	3 7%
	100.0%
By agency	
1 Schools and colleges	52%
2 Y M C.A.	26%
3 Other agencies	12%
4 Graduate study	6%
5 Commerce, industry, etc.	4%
	100%
3. Permanence of Field of Initial Placement (3)	
[illegible]	50%
[illegible]	68.5%
[illegible]	21 5%
[illegible]	10.0%
[illegible]	90%

REFERENCES

1. Arsenian, Seth *Comparison of Three Freshman Classes in Scholastic Aptitude, Achievement, Evaluative Attitude, Adjustment, and Interest* Springfield College, 1941. Mimeographed
2. Arsenian, Seth. *A Study of Student Mortality in Springfield College.* Springfield College, 1940. Unpublished
3 Domas, I J *A Study of Springfield College Graduates with the Bachelor's Degree, 1921–1940.* Springfield College, 1941. Unpublished Master's Thesis.
4 Frasher, C B *Freshman Week Programs.* Springfield College, 1936. Unpublished Master's Thesis
5. Hunsicker, Paul *A Mechanical Analysis of Chinning.* Springfield College, 1941 Unpublished Master's Thesis
6. Konstantinov, Krum *An Analysis of Some Soccer Skill Variables* Springfield College, 1940. Unpublished Master's Thesis

7. Larson, Leonard *Physical Tests and Measurements* Springfield College, 1941
8. Larson, Leonard. "A Factor and Validity Analysis of Strength Variables and Tests with a Test Combination of Chinning, Dipping, and Vertical Jump" *Research Quarterly,* XI (1940), Q-281.
9 Matthews, John *Predicting Academic Success of Freshmen.* Springfield College, 1941, Unpublished Master's Thesis.
10. Matthews, John *Factors Predicting English Test Scores and Freshman English Grades.* 1940, Springfield College. Unpublished
11. Quamme, H J. *Factor Analysis of Cardio-Vascular Tests and Variables.* Springfield College, 1941, Unpublished Master's Thesis.
12 Tibbetts, H N *The Development and Evaluation of Potential Basketball Ability Variables and Tests* Springfield College, 1940, Unpublished Master's Thesis.
13. Todd, J E *First Annual Report of the Common Freshman Year at Springfield College,* 1938. Unpublished
14. Todd, J. E *Annual Report of the Director of Admissions, Springfield College, 1939–40.* Unpublished
15 Todd, J. E *Social Norms and the Behavior of College Students* Teachers College. Contributions to Education No 833. New York. Bureau of Publications, Teachers College, Columbia University, 1941.
16. U S Office of Education *College Student Mortality* Bulletin, 1937. No. 11, Washington: U. S. Government Printing Office, 1938.

GUIDANCE THROUGH SELF-APPRAISAL

FORREST E. HEWITT
Howe Military School

It is well recognized that some guidance programs fail because the records become so elaborate and cumbersome that no one uses them. In the guidance program at Howe Military School, this difficulty has been overcome by arranging the material so that a student's problem can be ascertained with from three to five minutes of careful study. The cumulative record system is based upon the recognition of the following three problem fields.

(1) Self-guidance

If today's students are to be prepared to adjust themselves to this gadget-filled world of ours, it is most necessary that they be encouraged to self-guidance in the beginning of their high-school careers so that they can become more self-reliant, can make better choices, and will be better able to meet new situations wisely This awakening to the realization that one must be able to guide himself is one of the most important single factors of an individual's early experience.

(2) Professional adjustment

Too many high schools have allowed their seniors to leave school with the sole ambition of going to college if their parents could afford to send them. Many of these youngsters have gone to college with no further aim than to join certain fraternities or sororities. Education again was incidental and the professional careers were too far removed for immediate consideration. Too often these graduates have found that they took the wrong courses and would probably have been happier if they had chosen other fields of interest. These people too often

become professional misfits and spend the rest of their lives wishing that they had chosen other careers

(3) Self-appraisal

Students should and can be taught how to learn to appraise themselves, how to take advantage of opportunities for growth, and how to evaluate their progress. If a self-explanatory manual consisting of appraisal profiles and check sheets is assembled and written for the student needs of each particular school, the guidance program will not become lost in the daily factual teaching.

The source material for the Howe record system is recorded in profile form on a manila folder in which the cumulative material is filed This information is compiled from the following sources

(a) Technical information and personal inventory tests taken by the student.
(b) Reports from the Masters, Tactical Officers, the Headmaster, and the Commandant. (These reports give a cross section of the work habits, conduct, and social adjustments of the boy.)
(c) The Cadet's autobiography.
(d) Counselor's summary.

The Howe Profile System

During the four days of the orientation period prior to the fall term, a complete check is made on each of the entering student's achievements in the fundamentals of reading, language, mathematics, science, and the social studies. This battery is scored and placed on individual profiles showing the student's actual status in the tools of learning This information is sent to the Masters who begin their "special help" periods for the new student with specified difficulties

The entire program centers around the building of profiles showing achievements, interests, personality, and social adjustment. At various times during the school year all of the boys take a series of interest, personality, and special ability tests, the scores of which are placed on profiles by the Cadets for later

interpretation and analysis. Through the assistance of the counselor, each boy writes his *own appraisal* after drawing profiles of the various standardized tests

The first of the series of tests is a check on their technical information in the ten major fields, namely: human relations, literature, government and history, biological sciences, physical sciences, machine shop, mathematics, business administration, office techniques, and architectural art On this same profile the student's high-school training and occupational preferences are charted according to national norms for the academic level.

Each boy then takes a series of interest tests in the same ten major fields and compares these percentiles with his profiles in technical information, in occupational choices, and high-school training. These profiles are then compared with the vocational plans the student has outlined in his autobiography. Each boy writes a summary analysis of these profiles in the form of a letter or essay which is handed to the guidance counselor who interviews each boy in regard to his interpretation These letters are then given to the English Master who checks them for grammatical errors and letter-writing form. The corrected copies are then returned to the counselor who sends them to the boys' parents with personal letters of explanation suggesting discussions between parents and sons during the spring vacation concerning the personal and vocational implications as shown by the profiles.

At the first opportunity after his return from vacation, the Cadet adds to his cumulative record in the guidance office the additional information gained in his home discussions. If it is the boy's senior year, this information is used as a basis for information in his future life plan. If he plans to attend college the next year, this information becomes the basis for his Freshman college conferences and is placed in a folder and sent to the college of his choice.

Previous to his senior year, special aptitude tests are given to the Cadet and a complete inventory accumulated during the years he attends Howe School This personal inventory starts as soon as the student begins his formal life in the classrooms and the barracks, and accumulates from the following sources.

(1) Profiles of the boy's daily scholastic work which are checked by each Master under whom he is taking courses.
(2) Tactical Officers' reports showing a picture of the boy's adjustments to life in the barracks
(3) Headmaster's profiles showing ratings of respect for authority, leadership, ability and willingness to co-operate.
(4) Commandant's report summarizing punishments and reactions

Since the rating on many of these factors must of necessity be by subjective judgment, these profiles on the student's work habits, emotional control, cooperation, leadership qualities, and future plans are arrived at from various sources, thus giving a more reliable evaluation This summary profile in the form of a personal adjustment inventory has proved invaluable for counselling purposes and is placed in the Cadet's cumulative record file in the guidance office.

After the analysis of the profiles and inventories, there is no attempt to pigeon-hole the student into a given profession or particular college. Instead, wise counseling is combined with student freedom of choice The selection of professions is often born in the bi-monthly meetings of the College Preparatory Club For the under-classmen the broad choice of whether the boy wants to work with ideas, things, or people very often comes from his study of the career data and charts in the guidance office. This selection of a career by including the boy in the program from its inception gives the boy the feeling of responsibility and security, as he knows that he has some definite and specific plans for his future which can be changed, if necessary, to meet new situations.

Use of the Autobiography

In order to help the high-school student to know himself, the autobiographical method has proved indispensable. One senior entering the school for the first time last fall, stated at the close of his autobiography, "I still don't know exactly what I want to do, but writing this has helped me to know myself better, and I think by the close of the semester I will be able to make an intelligent decision."

During his Junior year, another boy decided that he wanted to become a chemical engineer, but when he and the counselor discussed his aptitude and achievement profiles, he was made aware of a definite weakness in mathematics (The author believes that one of the most effective features in this technique is that it offers the counselor the opportunity to assist the student in discovering his own interests and deficiencies.)

When writing his autobiography, this boy stated that his father was too busy to discuss his personal and vocational problems with him. His counselor suggested that if the boy presented a definite plan to his father, he was sure that his busy father would take time to advise him This Junior not only set up a definite program, but also followed the counselor's suggestions to try out different jobs during his vacation.

The broadening experience of this boy through the exploratory jobs, and the new rapport developed between father and son brought forth the proud statement from the boy when he returned to Howe for his Senior year, "Now I know my dad, and we decided that it would be better for me to change my career from engineering to dentistry." He went on to explain in his revised autobiography his reasons for changing his course, and gave in detail his new plans.

Autobiographies are written during the Freshman year and revised each consecutive year They are considered absolutely confidential and are filed with other cumulative material in each student's guidance folder in the counselor's office. Certain parts of the following autobiographical outline are revised and brought up to date from time to time.

I Life history
    (a) Ancestry
    (b) Description of home life
    (c) What the boy thinks the parents expect of him
II. Difficulties
    (a) School
    (b) Home
    (c) Social
    (d) Physical
    (e) Any other personal problems

III Achievements
    (a) Records, medals, etc.
    (b) Sports
    (c) Jobs held
    (d) Money earned and money saved
IV Interests, hobbies, and activities
    (a) Description of interest in hobbies
    (b) Amusements preferred
    (c) Reading interests
    (d) Travel
V. Emotional needs
    (a) Personal and social adjustment in school
    (b) Religious
VI Life plans
    (a) Professional Interests
    (b) Vocational experiences
    (c) Analysis of abilities in relationship to chosen field of occupation
    (d) College and technical school plans (Seniors)
VII Summary
    A short interpretation of the implications of the test results in comparison with the autobiography and future plans as previously made with parents

Each autobiography is read carefully by the counselor, and if it is incomplete or not well written, the paper must be revised and rewritten.

Practical Outcomes

This type of cumulative record offers the counselor both tangible and objective facts for interviewing each student. These records also help the student to decide upon more definite future plans, and offer an incentive for him to do much better school work. The profile system gives needed motivation for the superior as well as the average and poor students who have previously failed to work up to their ability levels These objective records also bring to the students the importance of personality development and its effect upon their social and school life.

In meeting the needs of each student as an individual by teaching him how to study, how to choose and prepare for a vocation, how to get along with others, and how to interpret his experiences which will enable him to solve present and future problems, the plan is helping the boys to adjust themselves to the rapidly changing world into which they are being graduated.

MEASUREMENT ABSTRACTS*

Borow, Henry "The Growth and Present Status of Occupational Testing." *Journal of Counseling Psychology* VIII (1944), 70–79
Tracing the development of psychological testing in this country, the author points out the close relationship between periods of socio-economic change and the rapid growth and refinement of measuring instruments and techniques. He indicates the principal men and organizations responsible for this progress which extends from World War I and the general mental ability test, through the depression era when job-placement testing became widespread, to the current wartime testing programs in the armed services and industry The work of the Personnel Service Division of the Pennsylvania State College Extension Services is cited to illustrate present-day trends in employment psychology While not infallible and admittedly subject to abuses, occupational testing is considered to have proved its worth and will play an important role in the postwar days to come. *Vernon S Tracht.*

Brown, Fred. "An Examination in 'Preventive' Testing in the Kindergarten." *Mental Hygiene*, XXVIII (1944), 450–455.
This describes an experimental program instituted in the Minneapolis public school system whereby kindergarten teachers are trained to administer the *Stanford-Binet* scale to their own pupils. The purpose is to better enable these teachers to recognize and prevent early indications of childhood maladjustment. Thus, when they can observe their children individually in the testing situation and later carefully analyze test responses item by item, they can use the *Binet* as an important diagnostic tool rather than only as an index of comparison between children. *Vernon S. Tracht.*

Cattell, Raymond. "A Note on Correlation Clusters and Cluster Search Methods." *Psychometrika*, IX (1944), 169–184.
Four methods of determining the clusters in a correlation matrix are described and compared. The choice of method has to be made according to the size of the matrix and the type of cluster sought. The relativity of clusters is emphasized and a distinction is drawn between phenomenal clusters and nuclear clusters. The relative utility of clusters and factors is briefly commented upon. (Courtesy *Psychometrika.*)

Davis, Frederick B. "Fundamental Factors of Comprehension in Reading" *Psychometrika*, IX (1944), 185–197
A survey of the literature was made to determine the skills involved in reading comprehension that are deemed most important by authorities. Multiple-choice test items were constructed to measure each of nine skills thus identified as basic. The intercorrelations of the nine skill scores were factored, each skill being weighted in the initial matrix roughly in proportion to its importance in reading comprehension, as judged by authorities. The principal components were rather readily interpretable in terms of the initial variables. Individual scores in components I and II are sufficiently reliable to warrant their use for practical purposes, and useful measures of other components could be provided by constructing the required number of additional items. The results also indicate need for workbooks to aid in improving students' use of basic reading skills. The study provides more detailed information regarding the skills measured by the *Cooperative Reading Comprehension Tests* than has heretofore been provided regarding the skills actually measured by any other widely used reading test. Statistical techniques for estimating the reliability coefficients of individual scores in principal-axis components, for determining whether component variances are greater than would be yielded by chance, and for calculating the significance of the differences between successive component variances are illustrated (Courtesy *Psychometrika.*)

---

* Edited by Forrest A. Kingsbury.

Grossman, David. "Technique for Weighting of Choices and Items on I B.M. Scoring Machines." *Psychometrika*, IX (1944), 101–105.
A technique has been developed which permits the weighting of responses of test items on the I B.M. scoring machine on the initial scoring, heretofore impossible. This is done by making the length of the response lines on the answer sheet longer or shorter as weights are needed. It is anticipated that this method will prove useful wherever differential weighting serves to increase the validity of tests. (Courtesy *Psychometrika.*)

Hastings, J T "Tensions and School Achievement Examinations." *Journal of Experimental Education*, XII (1944), 143–164
This article describes a technique for differentiating pupils on the basis of tension intensity in examination situations, the purpose being to discover possible relationships between their tension ratings and examination scores. In the course of a semester, ninth-grade students in mathematics at the University of Chicago High School were given four examinations graded according to importance, length, and difficulty A short questionnaire, which was designed to record their attitudes and feelings immediately following each, proved reliable by the split half method (mean coefficient of .76) and valid by its correspondence with the Luria technique which was administered twice prior to each examination. Predictability of examination results from tension scores was found feasible, but other data regarding relationships between these two variables were considered suggestive rather than conclusive. *Vernon S. Tracht*

Holzinger, Karl J "Factoring Test Scores and Implications for the Method of Averages." *Psychometrika*, IX (1944), 155–167.
The general procedure and detailed steps for attaining complete factor analyses of scores are presented. Both orthogonal and oblique factors are considered. It is shown that a single average by conventional procedure gives an incomplete summarization of the data when the rank exceeds one. There should be as many averages as there are common factors (Courtesy *Psychometrika.*)

Johnson, Palmer O. and Tsao, Fei. "Factorial Design in the Determination of Differential Limen Values." *Psychometrika* IX (1944), 107–144
This paper discusses the application of the principles of factorial design to an experiment in psychology For the purpose of illustrating the principles, a simple experiment was designed dealing with the determination of the differential limen values of subjects for weights increasing at constant rates. The factorial design was of the type 4 rates × 7 weights × 3 sexes × 2 sights × 2 dates. The appropriate statistical analysis for this type of design is the analysis of variance. The mathematical formulation of the problem was specified and the appropriate solution for the specific problem was obtained. Greater precision results from this type of design, in comparison with the traditional psychological experiment dealing with a single factor, in that measures are obtained of the effect of each of a number of factors together with their interactions. (Courtesy *Psychometrika.*)

Krathwohl, W C. "A 3 by 3 Analysis of the Predictive Value of Test Scores." *Journal of Applied Psychology*, XXVIII (1944), 318–322.
Because correlations between educational test scores and academic achievement are usually too low to be used for predictive purposes, another method of determining the relation between achievement and ability has been developed. Percentile scores on a general ability test were divided into three groups, (1) above the 75th percentile, (2) between the 25th and 75th percentiles, (3) below the 25th percentile. Achievement grades were also divided into three groups. When the test scores and

grades were tabulated in a 3 by 3 table, information useful in counselling and teaching practices as well as in prediction of success in academic work was obtained. *L. Bouthilet.*

Lewinski, Robert J, Lt. Comdr, H-V (S), USNR "Discriminative Value of the Sub-Tests of the Bellevue Verbal Scale in the Examination of Naval Recruits." *Journal of General Psychology*, XXXI (1944), 95-99

The original form of the *Bellevue Verbal Scale* was administered to 451 naval recruits at the Great Lakes Naval Training Station who were referred for examination to determine their fitness for military duty In terms of the mean weighted scores the performances of each of the four I.Q. categories on each of the five sub-tests (arranged in order from best to worst) were as follows Normal Group (I.Q. 91-110; Mean I.Q. 100.22, N 51): Comprehension, Information, Arithmetic, Similarities, Digit Span, Dull Normal Group (I.Q. 80-90, Mean I.Q. 83.73; N 53) Comprehension, Digit Span, Arithmetic, Similarities, Information; Borderline Group (I.Q. 66-79; Mean I.Q. 71.73, N 189) Comprehension, Digit Span, Similarities, Information, Arithmetic, Mentally Deficient Group (I.Q. below 66, Mean I.Q. 59.16; N 158): Comprehension, Similarities, Information, Digit Span, Arithmetic. The standard deviations of the various groups for the total scale, and to a lesser degree for the sub-tests, were relative to the I.Q. ranges. All critical ratios were sufficiently large to indicate that the obtained differences are statistically significant. The data reported indicate that all the sub-tests are satisfactory in differentiating the four categories. *Marion H Groves*

Lord, Frederic M. "Reliability of Multiple-Choice Tests as a Function of Choices per Item." *Journal of Educational Psychology*, XXXV (1944), 175-180.

Changing the number of choices per item in a multiple-choice test changes the reliability of the test. In order to express the amount of change a formula based on the Spearman-Brown prophecy formula is derived. The reliability of the revised test, with a change in number of choices per item, is given in terms of the reliability of the original test, and a constant value is determined by the percentage of individuals answering the items correctly and by the number of choices per item in the original and revised tests. Although an empirical check demonstrated the usefulness of the formula, further experimentation is to be desired. *L. Bouthilet.*

Manson, Morse P "The Concepts of the Profile, Psychograph, and Evalograph." *Journal of Educational Psychology*, XXXV (1944), 145-156.

Over a hundred different terms have been used to describe the graphic devices of the profile, the psychograph (both in use for at least 30 years), and the evalograph (developed here for the first time) Precise definitions are needed Taking into consideration some of the literature on the subject the following definitions are offered (a) A "profile" is a graphic curve or line constructed from many individual or group measurements and/or estimates expressed in identical terms. (b) A "psychograph" is a condensed analytical record of a complex of factors, containing a profile and related numerical and descriptive data (c) An "evalograph" is a psychograph with additional interpretive and control data, requiring clinical techniques for most effective use. *Marion H Groves*

Older, H J "An Objective Test of Vocational Interests." *Journal of Applied Psychology*, XXVIII (1944), 99-108.

The author presents the method and results of a new technique in measuring vocational interests. Pictures representing various occupational activities were projected onto a screen, after seeing which the subjects—336 high-school students, 248 college students, and 59 business college students—were tested on the material they had viewed. Scores for those who took supplementary tests also revealed low correlation with intelligence scores and with the *Strong Vocational Interests Blank*. However, close relationship was found between expressed ambitions and test results. This vocational interest test is regarded as an objective measure of an individual's dynamic interests in contrast with the subjective quality of the Strong Inventory *Vernon S. Tracht.*

Poston, F "An Evaluation of Word and Picture Tests for First and Second Grades." *Journal of Applied Psychology*, XXVIII (1944), 142-152.

Three widely used primary-grade word recognition tests were presented individually to 100 unselected first- and second-grade pupils equated for age, M.A., sex, school grade, and achievement in reading. Of four methods of presentation, the children obtained highest scores when asked to select a printed word after hearing it pronounced, and when identifying a word when shown a corresponding picture. Considerably more difficult were pronouncing a word without a picture and giving the "correct" name for pictures alone. Analyses of the items revealed that the words in the tests are not graded in difficulty and that a great amount of variability is present in responses to the same pictures. The presence of the pictures was sometimes a retarding factor because they suggested to the children words other than those intended by the authors of the tests. The same words appearing with different distracters became more difficult as the distracters approached the form or meaning of the test item itself. *Edith S. Jay*

Rashevsky, N. "Contributions to the Mathematical Theory of Human Relations, VIII: Size Distribution of Cities" *Psychometrika*, IX (1944), 201-215.

An attempt is made to connect the distribution function of the sizes of the cities with the distribution functions of some other characteristics of the individuals in the society Several theoretical possibilities are discussed and different relations are derived. A possible connection with some observed relations is discussed. (Courtesy *Psychometrika*.)

Rautman, A. L. "Performance of Mental Defectives on the Revised S-B and the Kent E-G-Y Tests." *Journal of Applied Psychology*, XXVIII (1944), 329-333.

Three hundred and fifty mentally retarded patients were given the *Kent Oral Emergency Test* and then the *Revised Stanford-Binet* in order to determine the relationship between the two. Scores showed the *Kent* to be approximately 10 I.Q. points higher, with a coefficient of correlation between the I.Q.'s of .84±.01 An analysis for internal consistency revealed several unsuitable questions. When allowance is made for the difference in I.Q. found in this study, the *Kent E-G-Y*, which contains twenty five chosen questions and takes only ten minutes to administer, can be used to advantage in clinical situations where a short and simple, yet fairly accurate, preliminary test of mental ability is required before any corrective steps are undertaken. *Vernon S, Tracht.*

Rosenzweig, Saul; Bundas, Lourene E., Lumry, Kelly; and Davidson, Helen. "An Elementary Syllabus of Psychological Tests." *Journal of Psychology*, XVIII (1944), 9-37

An explanation is presented of the purpose, materials, instructions, data, scoring methods, and interpretation of five types of psychological tests with representative tests of each type. A. Tests of Intelligence: *Stanford-Binet, Cattell Infant Intelligence Scale, Wechsler-Bellevue Intelligence Scale, Otis Self Administering Test of Mental Ability, A Point Scale of Performance Tests, Form I* B Tests of Intellectual Deterioration *Babcock Levy Revised Examination for the Measurement of Efficiency of Mental Functions, Goldstein-Scheerer Cube Test of Abstract and Concrete Behavior, Wells Memory Test* C. Tests of Vocational Aptitude *Minnesota Mechanical (Aptitude) Assembly Test, Seashore Measures of Musical Talents, Strong Vocational Interest Inventory* D. Tests of Personality *Bernreuter Personality Inventory, Allport-Vernon Study of Values, Word Association Test.* E. Projective Approaches to Psychodynamics: *Rorschach Test, Thematic Apperception Test, World Test, Rosenzweig P-F (Picture-Frustration) Test* *Betty Steele*

Rusmore, J T "Comparison of an 'Industrial' Problem-Solving Task and an Assembly Task." *Journal of Applied Psychology*, XXVIII (1944), 129-131

The *Crawford Test of Tridimensional Structural Visualization* was administered ten times in succession to 87 women and 44 men university students. Assuming that the first trial measured tridimensional structural visualization, while the tenth trial measured an "assembly function," the two groups were compared on each of these two abilities. The critical ratios for the ten trials varied from .03 to .37, indicating no

significant sex differences in mean levels of performance on any trial. The first and tenth trial for the women correlated .36±.12; for the men, .09±.15. The university students differed slightly from the original Crawford standardization group in initial performance, but the restriction in the range of talent is not sufficient to explain the low correlations obtained. It is concluded that no sex differences in mean levels of performance were found for either function. *Edith S Jay*

Spache, George. "The Abbreviated Stanford-Binet Scale in a Superior Population." *Journal of Educational Psychology* XXXV (1944), 314-318.

Wright's modified short scale of the *Stanford-Binet* was administered to one hundred pupils of a nursery-school and two private schools in New York City. I.Q.'s below average were not used. It was found that the modified scale was superior to the clinically used short scale in range, size, and frequency of errors for predicting long-scale results. The modified scale also had greater predictive value in a subnormal population than in a superior population. However, the relatively large prediction error makes the use of the full scale preferable. *Betty Steele.*

Sumner, F C. and Brooker, N M "Prognostic and Other Values of Daily Tests." *Journal of Applied Psychology*, XXVIII (1944), 323-328.

The records of students who were given short, objective, daily tests in a course in General Psychology were analyzed. The correlation between the averages on the first 5 tests and on all 40 tests was .82 Although the mean score of all students increased, the poorer students showed a greater increment. It is concluded that the prognostic value of daily tests is more reliable as an indicator of success in the course than that of general intelligence test results. Moreover, the program is of value in the learning and motivational aspects of the situation. *L. Bouthilet.*

Thurstone, L. L. "Second-Order Factors" *Psychometrika*, IX (1944), 71-100.

Second-order factors are defined and illustrated in terms of a literal notation, a physical example, a diagrammatic representation, a geometrical example, and the matrix equations relating the first-order and second-order domains. Both kinds of factors are discussed as parameters which may be not only descriptive of the individual objects in a statistical population but also descriptive of the restrictive conditions under which the objects were generated or selected. Second-order factors may be of significance in reconciling the several theories of intelligence. This paper is concerned with test configurations that show simple structure. If such a structure is not revealed, then the second-order domain is indeterminate. (Courtesy *Psychometrika*.)

Tsao, F "A Study of the Relationship between Grade and Age Variability" *Journal of Experimental Education*, III (1944), 187-200.

The *L. Test* and Bartlett's procedure were applied to data gathered by Thorndike, Whitmire, and Pressey. The results indicated that the difference in variability from age to age is not significant. The *L. Test* and Bartlett's procedure then were applied to two Canadian public schools, grades five and eight. Materials used were the *National Intelligence Test, Scale A, Form 1, Schorling-Clark-Potter Arithmetic Test, Form A*, and the *Otis Reading Survey, Form I.* Results indicated that variability is constant from grade to grade and age to age. The relationships among mental and scholastic functions are positive but not perfect, and these relationships are constant from grade to grade. On the basis of these results, it is suggested that individual differences should be emphasized as early as possible by the placement of a child in a subject according to his ability All children in the same class would then have approximately the same standing. *Betty Steele.*

Tucker, Ledyard R. "The Determination of Successive Principal Components without Computation of Tables of Residual Correlation Coefficients." *Psychometrika*, IX (1944), 149-153.

A procedure is presented for determining the successive principal components of a correlation matrix where it is not necessary to compute the successive tables of residual correlations. The original correlation matrix is bordered with a new row and column for each principal component that is determined. (Courtesy *Psychometrika*.)

NEWS NOTES*

The A.C.P.A. has sent to each of its members copies of "Personnel Work in the Post War College and University," by J G Darley, and the American Council on Education brochure "Counseling and Post War Educational Opportunities," edited under the chairmanship of E. G. Williamson, of the Committee on Student Personnel work of the A.C.E.

As an organization, the A.C.P.A. endorses the proposal for an International Office for Education, to aid in the creation of more healthful relations among people and, through them, among nations of the world.

Thelma Mills, Chairman of the Membership Committee is pleased to introduce the following new members:

William M. Adams, 2nd Lt. AUS, 500 Brook Avenue, Lafayette, Louisiana.
Ruth L. Alexander, 804½ E. Eighth Street, Bloomington, Indiana.
Thelma Albright, Dean of Students, Queens College, Charlotte, North Carolina.
Gordon V Anderson, Acting Director, Student Counseling Bureau, University of Minnesota, Minneapolis, Minnesota.
Irwin A. Berg, Clinical Counselor and Associate in Psychology, University of Illinois, Urbana, Illinois.
Mrs. Dorothy Brooks, Dean of Women, Denison University, Granville, Ohio.
Paul J Brouwer, Cooperative Study in General Education, 5835 Kimbark Avenue, Chicago, Illinois.
John W Bunn, Dean of Men, Stanford University, California.
Lt. John Butler, A.S.T.P., SCU 3700, University of Minnesota, Minneapolis, Minnesota.
Dean C. W Cannon, Cooperative Study in General Education, 6010 Dorchester Avenue, Chicago, Illinois.
O. R. Chambers, Head, Department of Psychology, Oregon State College, Corvallis, Oregon.
Barbara M Clark, Assistant Director, Student Activities Bureau, University of Minnesota, Minneapolis, Minnesota.
Harold W Colvin, Y.M.C.A. National Council, 19 S. La Salle Street, Chicago, Illinois.
James H. Corson, Dean of Men, College of the Pacific, Stockton, California.
Margaret L. Cunningham, Dean of Women, Ripon College, Ripon, Wisconsin.
Marguerite R Dressler, Assistant Director of Personnel, Florida State College for Women, Tallahassee, Florida.
Charles V Dunham, Director of Student Employment, University of Texas, Austin, Texas.
Wilma Eldersveld, Psychological Clinic, University of Michigan, Ann Arbor, Michigan
Mary Catherine Evans, Vocational Adviser for Women, Indiana University, Bloomington, Indiana.
Marian Farnes, Assistant Dean of Students, College of City of New York, New York, N Y
Mrs. Leona Wise Felsted, Chairman, Personnel Committee, Illinois Wesleyan University, Bloomington, Illinois.
Helen E. Focht, Assistant Director, Office of Student Affairs, State University of Iowa, Iowa City, Iowa

---

* News items concerning members of the American College Personnel Association should be sent to Grace E. Manson, Northwestern University, Evanston, Illinois.

John D Foley, Assistant to the Dean, Office of Dean of Students, University of Minnesota, Minneapolis, Minnesota.
Mary Lou Griswold, Placement Secretary, New York State College of Home Economics, Cornell University, Ithaca, New York.
Thomas J Hand, Vocational Advisory Service, 87 Madison Avenue, New York, New York.
Dr W Ballentine Henley, President, College of Osteopathic Physicians, Los Angeles, California
George Hilliard, Director of Personnel and Guidance, Western Michigan College of Education, Kalamazoo, Michigan.
George E. Hill, Director, Student Personnel Service, Macalester College, St. Paul, Minnesota
Arthur C. Hicks, Acting Registrar, Western Washington College of Education, Bellingham, Washington
M Eunice Hilton, Dean of Women, Syracuse University, Syracuse, New York.
Lela L. Holcomb, Assistant Counselor, University Residences for Women, University of Texas, Austin, Texas.
Wilbur J Humber, Dean of Student Affairs, Kalamazoo College, Kalamazoo, Michigan.
Katherine J Jones, Executive Secretary, Nursery Training School of Boston, 355 Marlborough Street, Boston, Massachusetts.
Flemmie P Kittrell, Dean of Women, Hampton Institute, Hampton, Virginia.
Mrs. Estella Hitchcock Lane, Dean of Students, Frances Shimer College, Mt. Carroll, Illinois.
Corinne G. La Barre, Research Assistant, Western Personnel Service, Pasadena, California
Frances A. Lahr, Director, Student Union, University of Nebraska, Lincoln, Nebraska.
Howard C. Madsen, Manager Technical Employment and Training, Westinghouse Electric and Manufacturing Company, Pittsburgh, Pennsylvania.
Betty Anne Malloy, Director of Appointment Bureau, Manhattanville College of the Sacred Heart, New York, N Y
Thomas O Marshall, Director of Student Personnel, Colorado State College, Fort Collins, Colorado.
Harry M Mason, Testing and Guidance Program, University of Texas, Austin, Texas.
Catharine Northrup, Assistant Director, Student Personnel, Colorado State College, Fort Collins, Colorado.
Dorothy Olsen, Director, Texas Union, University of Texas, Austin, Texas.
L. R. Palmerton, Director Student Personnel, South Dakota School of Mines, Rapid City, South Dakota
Penelope M. Pearl, Psychological Clinic, University of Michigan, Ann Arbor, Michigan.
C R Penoi, Faculty Exchange, University of Oklahoma, Norman, Oklahoma
Harold B. Pepinsky, Office of Dean of Students, University of Minnesota, Minneapolis, Minnesota.
Julia E. Read, Personnel Director, College of St. Elizabeth, Convent Station, New Jersey
Thomas F Richardson, Director Student Personnel, Texas Christian University, Fort Worth, Texas.
Howard L. Roy, 507 Oak Street S.E., Minneapolis, Minnesota
Chester H Ruedisili, Assistant Dean of College of Letters and Science, University of Wisconsin, Madison, Wisconsin
John Dale Russell, Dean of Students, Division of Social Science, University of Chicago, Chicago, Illinois
William H Seaman, Director of Admissions and Bureau of Appointments, Oberlin College, Oberlin, Ohio.
Jane E. Stephenson, Acting Placement Director, University of Detroit, Detroit, Michigan
Mrs. Jean T Stromberg, Assistant to the Counselor, Cornell University, Ithaca, New York.

C. Woody Thompson, Director, Office of Student Affairs, University of Iowa, Iowa City, Iowa.
Alice L. Thorpe, Appointment Office, Wheaton College, Norton, Massachusetts.
M. R Trabue, Dean, School of Education, Pennsylvania State College, State College, Pennsylvania
Leona E. Tyler, Director, Bureau of Personnel Research, University of Oregon, Eugene, Oregon.
Mrs. Frances Gibson Wallace, Freshman Adviser, Hollins College, Hollins, Virginia.
Robert S Waldrop, Clinical Assistant, Psychological Clinic, University of Michigan, Ann Arbor, Michigan.
Mrs. Dorothy S. Whitehead, Supervisor of Student Aid, University of Georgia, Athens, Georgia
Jane Louise Winningham, University Elementary School, Ann Arbor, Michigan.
Austin B Wood, Assistant Professor of Psychology, Brooklyn College, Brooklyn, New York.

Elizabeth G Andrews, Director of Personnel, Florida State College for Women, has prepared a manual for the use of faculty counselors at her institution.

A. J Brumbaugh has recently accepted the Vice-Presidency of the American Council on Education, Washington, D C.

Harold B Pepinsky has been appointed Counselor, Vocational Counseling and Guidance Center, University of Kansas, Lawrence, Kansas.

Lieut. E. L. Swann, U.S.N R., has been transferred from the U. S. Naval Flight Preparatory School at the University of South Carolina to the Naval Air Training Center, Corpus Christi, Texas.

TRIAL AT YALE UNIVERSITY OF THE ARMED FORCES INSTITUTE GENERAL EDUCATIONAL DEVELOPMENT TESTS

ALBERT B. CRAWFORD
AND
PAUL S BURNHAM
Yale University

Among several major problems relative to demobilization with which our colleges are immediately confronted is that of evaluating, fairly and with reasonable assurance, the scholastic promise of individuals now serving the Armed Forces Educational provisions of the "G.I Bill" in due course undoubtedly will bring to the colleges a flood of applicants for admission. Among these will be many "irregulars"—those with broken or deferred educational histories in the formal sense, yet with valuable training or new skills acquired in service Since these personal developments will be difficult, if not impossible, to assay in traditional coinage of the academic world ("units" and "credits"), colleges at last may be impelled to consider what the prospective student *knows,* and what educational aptitudes he can *demonstrate,* regardless of where or how these attributes have been obtained.

The Armed Forces Institute has made educational programs of extensive scope available throughout all major theatres of training and operation Certificates attesting satisfactory completion of particular courses, or achievement-test scores in specific fields, are issued through "G.H.Q." for this purpose, at the University of Wisconsin, Madison, However, the Armed Forces Institute measure most useful to colleges in their selection of future, demobilized students appears to be the *General Educational Development Battery* of four tests, viz·

Test 1) *Correctness and Effectiveness of Expression*
Test 2) *Interpretation of Reading Materials in the Social Sciences*
Test 3) *Interpretation of Reading Materials in the Natural Sciences*
Test 4) *Interpretation of Literary Materials.*

No further discussion of these tests is necessary here, since they have been widely publicized and are fully described in the Examiners' Manual issued by the American Council on Education.[1] This Manual contains valuable remarks concerning the nature and purposes of the *General Educational Development Battery*, and presents tentative college norms for each test. It stresses the need for local norms as well, stating.

> These norms are intended to help the schools decide what minimum test performance should entitle returning war-service persons to a given amount of academic credit in a given area. Because of the heterogeneity just noted, however, there is danger that certain schools, by uncritical acceptance of the general norms here reported, may set standards inconsistent with their local needs It is highly desirable, therefore, that each institution, by administering the civilian forms to its own civilian students establish its own local norms for use along with the general norms in the interpretation of scores of service man and women reported to the institution by the United States Armed Forces Institute.

To that end, particularly as it might affect the program of Yale Studies for Returning Service Men, an experiment was recently conducted at Yale University All members of the civilian freshman class which had matriculated in July 1944 were invited to participate; 135, or about one-third of the total, did so. Although the invitation put no pressure upon freshmen—since only interested volunteers were desired—it stated: a) that participation would be of direct service to the University and collaterally to men now in the Armed Forces, b) that reasonable compensation would be given for the time expended,

---

[1] For further information on the Battery and other phases of the Armed Forces Institute so-called "Fox-hole University" see *The United States Armed Forces Institute, Tests of General Educational Development (College Level) Examiners' Manual,* American Council on Education, 1944; and *Guide to the Evaluation of Educational Experiences in the Armed Services,* American Council on Education, Washington, D. C., 1944.

For each of these four tests comprising the A.F.I *General Educational Development Battery*[a] two hours are allowed.

In appraising their performance, it is essential to note how well the test group represented, in scholastic promise and achievement, the entire class. Fortunately it was found that the random sample of participants in this experiment covered virtually the entire class range, from highest to lowest, in academic standing, for the first complete term of Freshman Year. Average score of this group on the College Entrance Examination Board's *Scholastic Aptitude Test* was 575 and on other tests of the College Board or the *Yale Educational Aptitude Battery* of the same order—around .20 standard deviation above the class mean Hence these voluntary participants ranked, in pre-matriculation indices and later scholastic achievement somewhat above the median of Yale freshmen; i e, around percentile 60 rather than 50 in academic promise and accomplishment. Moreover, despite relatively high average performance of this freshman group on the A F.I tests, *dispersion* of their scores followed a normal-probability pattern

TABLE 1

AFI Test Group Compared with Total Freshman Class of 1947

	Mean	S.D
CEEB Verbal Scores		
[illegible]	555	98
[illegible]	575	101
[illegible]		
[illegible]	583	80
[illegible]	595	84
[illegible]		
[illegible]	73.9	7.4
[illegible]	74.6	7.1
[illegible]		
[illegible]	75.3	8.4
[illegible]	76.5	7.8

* Irregularities in secondary-school record obviated scholastic predictions in some cases; elimination, chiefly for military service, among freshman matriculants naturally accounts for the difference between 433 and 411 in total class numbers represented above.

---

[a] The Battery employed is that obtainable through the Cooperative Test Service, 15 Amsterdam Avenue, New York City—a parallel form to that actually used by the Armed Forces Institute.

Time Allowances

Maximum time allowed for each test is two hours As indicated in Table 2, many Yale students finished in considerably less time. The low correlations there shown also indicate that length of time spent on each test bore little relationship to score made by these freshmen. The time allowances are generous and the tests (as intended) therefore seem to depend on "power" rather than on speed

TABLE 2

Correlation of AFI Test Scores with Testing Time for Yale Freshman Test Group (N=135)

	Average time	r with time
AFI Test I (Expression)	65 min.	−.12
AFI Test IV (Literature)	75 min.	.15
AFI Test II (Social Studies)	80 min.	−.01
AFI Test III (Natural Science)	93 min.	.04

Intercorrelations among the Armed Forces Institute Tests

Since all four tests of this battery, including that designated as "*Interpretation of Reading Materials in the Natural Sciences*" are highly verbal, a considerable degree of positive intercorrelation among them would be expected The *Examiners' Manual* rather surprisingly offers no data in this respect. The following table presents these, for the Yale freshman test group. The correlations in this and subsequent tables were obtained from standard scores for the AFI tests, derived from raw scores by use of the conversion tables furnished in the *Examiners' Manual*.

TABLE 3

Intercorrelations among AFI Tests (N=135)

Tests	IV (Literature)	II (Social Studies)	III (Natural Science)
I (Expression)	.57	.47	.48
IV (Literature)		.63	.52
II (Social Studies)			.57

This correlation matrix indicates rather clearly (on the basis of a small but representative Yale sample) that the

Armed Forces Institute *General Educational Development* battery is general, rather than specific or differential, throughout its several parts. We hasten to add that no derogation is implied by the foregoing statement. It is simply not in the nature of this Battery, nor in the aims of its distinguished constructors, to measure *differential* rather than general scholastic aptitudes; which indeed should be clear from its title and from objectives discussed in the *Examiners' Manual* Nevertheless, these coefficients are not so high as to suggest that all four tests measure the *same* mental functions. It may be that similarity in *form* (i e., the fact that all four are highly verbal) accounts for the intercorrelations, and that dissimilarity in *content* accounts for their not being even higher

Comparison of Armed Forces Institute with College Entrance Examination Board Tests

Although the content and objectives of A.F.I. *General Educational Development Tests* differ considerably from those of College Entrance Examination Board measures, certain elements of the two series, from Yale evidence, appear functionally to have much in common. A chief difference of objectives concerns recency of formal schooling; i.e, College Board examinations are meant to be taken in stride by students nearing completion of their college-preparatory work and therefore to a substantial degree quite properly measure *current*, specific achievements. The A.F I educational development battery, no less properly, attempts to measure combined aptitude and achievement (however acquired) in more general terms. Yet the following table indicates a rather high degree of correspondence among those parts of each test-series which by title seemingly represent analogous educational fields. The Yale experiment affords direct comparison in this respect of the two testing methods. Candidates for admission to Yale are normally *required to take* the College Board *Scholastic Aptitude Test* (both verbal and mathematical sections) and the *English Essay Examination*. They distribute themselves rather widely among the other College Board options, according to subjects of study in the senior-high-school year and intended college program. Consequently some of the correlations represented below, and others in Table 3, are based upon too small a number of cases to warrant general conclusions By way of comparison, first-term versus second-term grades in the Freshman Year (Class of 1945W) correlated: English 10, .61, History 10, 73, Chemistry 11 and 14, .57; Physics 10, 76.

Despite the numerical limitations of these data in some categories, it appears evident that the Armed Forces Institute tests correlate surprisingly well with those of the College Entrance Examination Board in related areas—especially when allowance is made for the high average performance and therefore somewhat reduced "spread" of Yale freshmen on the A.F.I. Battery.

TABLE 4

Correlations between CEEB and AFI Tests

CEEB Tests	N	AFI Tests I (Expression) r	 IV (Literature) r	 II (Social Studies) r	 III (Natural Science) r
SAT (Verbal)	135	.63	.74	.64	.56
English Essay	135	.53	.50		
Social Studies	34			.71	
Chemistry	54				.77
Physics	58				.42

NOTE Average of all CEEB Tests correlated .72 with AFI Total score (N=135); for the means and standard deviations associated with the data of Table 4, the reader is referred to Table 4a at the end of this article.

Relationship of Scores on A.F.I. General Educational Development Battery and Certain College Board Tests to First-Term Grades at Yale

A pragmatic and customary, though by no means ideal, method of evaluating tests is to correlate scores thereon with subsequent grades in course. The latter, however carefully assigned, because of their subjective nature have dubious statistical reliability. However, in many situations they offer the fairest criterion by which the relative value of different prognostic measures may be compared. The next exhibit

(Table 5) presents a series of correlations between test scores and first-term grades, of participants in this experiment.

It is recognized that several factors bearing upon comparison of the College Entrance Examination Board tests and U.S.A.F.I *General Educational Development Tests* in Tables 4 and 5 may affect the results there given. All 135 men took the same A.F I tests but all did not take the same College Board tests. They all did take three of the latter, *Scholastic Aptitude Test (Verbal), Mathematical Aptitude* and *English Essay*. The remaining two College Board tests taken by each candidate for admission are elections determined by his secondary school course and prospective college major and are drawn from the following options: *Social Studies, French, German, Spanish,* or *Latin Reading; Biology, Chemistry,* or *Physics; Spatial Relations (Three-Dimensional) Visualising* Consequently the average on College Board Examinations is not based upon an identical battery for all men represented in this study. Each candidate's options, however, are appropriate to his expected freshman program.

TABLE 5

Correlation of CEEB and AFI Tests with Freshman-Year, First-Term Grades

(Data based on AFI Test Group of 135 Freshmen)

Course	Test	r	N
English 10a	AFI I (Expression)	.50	100
	AFI IV (Literature)	.54	100
	CEEB SAT	.53	100
	CEEB English Essay	.36	99
History (All Courses)	AFI II (Social Studies)	.52	43
	[illegible]	4[illegible]	4[illegible]
Physics 11–12a	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]
Mathematics 12a	AFI I (Expression)	.46	78
	AFI IV (Literature)	.30	78
	AFI II (Social Studies)	.47	78
	AFI III (Natural Science)	.30	78
	CEEB MAT	.59	78
Eng. Drawing 10a	CEEB Spatial	.63	35
Freshman First-Term Average	AFI I (Expression)	.51	135
	AFI IV (Literature)	.41	135
	AFI II (Social Studies)	.50	135
	AFI III (Natural Science)	.36	135
	AFI Total Score	.56	135
	CEEB SAT	.41	135
	*CEEB Verbal Average	.40	135
	†CEEB General Average	.44	135
	Average of all CEEB tests	.53	135

* *Average of CEEB Language, Social Studies and English Essay tests.*

† Average of all College Board achievement tests.

NOTE For the means and standard deviations associated with the data of Table 5, the reader is referred to Table 5a at the end of this article.

Most of the students represented took their College Board tests in April, 1944, whereas the A F I. battery was administered near the end of the first term. Consequently, the interval between taking College Board tests and obtaining the final grades for the term (six months) was considerably greater than that between A F I. tests and the end of term (one month) There is no way of estimating to what extent these differing time intervals may have affected correlations of the respective tests scores with the criteria.

Again the number of cases upon which some of these correlations are based permits only tentative conclusions. It does appear, however, that

1) A.F.I. tests show promise of being wholly acceptable alternates for College Board examinations in the verbal subjects.

2) For this sample group, A.F I Total Score correlated as well with Freshman first-term average in all courses, as did the average of all College Board tests.

3) The A F.I *General Educational Development Battery* makes no pretense of measuring abilities in Mathematics, Mechanical Drawing, or Descriptive Geometry. The College Board *M.A T.* and *Spatial Relations* tests are probably indispensable for prospective scientific or engineering majors. We have obtained no information regarding A.F.I. tests in specific branches of mathematics, but these would have pertinence chiefly for students who have completed those particular courses.

4) Since College Board *M.A.T* scores are likewise somewhat dependent upon previous mathematical training, interruption thereof might adversely affect individual performance *on this test also. By its nature,* the *Spatial Relations score is* less likely to be so affected.

From joint inspection of the A.F I. test scores and freshman grade distributions, it was a relatively simple matter to determine two critical-score levels Practically no students scoring above the upper one had unsatisfactory first-term records, while most of those scoring below the other ranked well under the class average. As a result of this investigation it was voted by the Executive Committee of Yale Studies for Returning Service Men to admit candidates scoring above the higher critical level on the A F.I. Battery despite their deficiency in formal academic credits, to reject or discourage (unless additional evidence more favorable to their chances of success is submitted) those scoring below the lower critical level, and to examine by other means border-line cases falling between these upper and lower levels.

For the reasons noted above in Conclusion (3) further tests (of the aptitude rather than formal achievement type) will be required of prospective engineering, mathematics or physical science majors.

TABLE 4a

Means and Standard Deviations of Table 4 Data

Variables correlated	N	Mean	S.D.
a) CEEB SAT and English Essay with AFI tests			
CEEB SAT	135	57.5	10.1
CEEB English Essay	135	55.5	8.5
AFI I	135	62.9	6.2
AFI IV	135	64.7	6.1
AFI II	135	69.5	7.7
AFI III	135	73.3	5.5
b) CEEB Social Studies and AFI II			
CEEB Social Studies	34	59.6	9.2
AFI II	34	69.3	8.1
c) CEEB Chemistry and AFI III			
CEEB Chemistry	54	56.2	8.8
AFI III	54	72.8	6.1
d) CEEB Physics and AFI III			
CEEB Physics	58	58.5	7.9
AFI III	58	74.3	4.4
e) CEEB average and AFI total score			
Average of all CEEB tests	135	57.4	6.3
AFI total score	135	270.4	20.9

TABLE 5a

Means and Standard Deviations of Table 5 Data

Variables correlated	N	Mean	S.D.
a) English Grades, AFI and CEEB tests			
English Grades (10a)	100	74.4	7.7
AFI I	100	62.0	5.9
AFI IV	100	63.8	5.9
CEEB SAT	100	55.0	8.8
English Grades (10a)	99	74.4	7.7
CEEB English Essay	99	54.3	8.5
b) History Grades, AFI and CEEB tests			
History Grades (all courses)	43	77.2	8.2
AFI II	43	69.4	7.6
CEEB SAT	43	55.8	9.9
c) Physics Grades, AFI and CEEB tests			
Physics Grades (11–12a)	44	78.4	12.2
AFI III	44	74.7	4.6
Physics Grades (11–12a)	28	78.6	11.5
C[illegible]	29	60.4	9.1
d [illegible]			
[illegible]	78	73.7	13.3
AF[illegible]	78	63.0	6.2
AFI IV	78	64.4	6.4
AFI II	78	69.1	7.9
AFI III	78	74.1	5.1
CEEB MAT	78	60.0	9.4
e) Engineering Drawing Grades and CEEB scores			
Engineering Drawing Grades (10a)	35	81.5	7.8
CEEB Spatial	35	57.5	9.5
f) Freshman Averages and CEEB tests			
Freshman Average (First Term)	135	76.5	7.8
AFI I	135	62.9	6.2
AFI IV	135	64.7	6.1
AFI II	135	69.3	7.7
AFI III	135	73.3	5.5
AFI total score	135	270.4	20.9
CEEB SAT	135	57.5	10.1
CEEB Verbal Average	135	55.4	7.4
CEEB General Average	135	56.2	6.3
Average of all CEEB tests	135	57.4	6.3

THE CRITERION

HERBERT A. TOOPS
The Ohio State University

In all test work, and in making predictions generally, there must be—according to the current mode of the thought and its consequent development into corresponding formulas—a *unitary, general* success score, or criterion score, for each person of the experimental group by whose aid the tests are constructed, combined, and validated

Item analysis for selecting the better items requires such a criterion. So does item alternative analysis, done with the purpose of altering, in the hope of improving, those confusion alternatives which are "out of line," which have, for example, positive item-alternative validity coefficients instead of the much-to-be-desired negative ones. For a multiple-choice test we desire such alternatives—for right answer and confusion alternatives—that the validity of each item may be as high as possible; or, to put it in another fashion, that as many as possible of the experimental items shall have a validity coefficient above some arbitrary lower limit, say .24 for tests of intelligence at the university freshmen level In the former case, which presupposes the second having already been satisfactorily accomplished, we are concerned with the picking of the individual items such that, for the *n* items chosen for a given test, the correlation of the sum of the scores (the Rights score) on the items-as-a-whole (i.e., the test as finally constituted) and the criterion shall be a maximum If the intercorrelations of the items are low, approximating zero, we may be content to ignore them; but we can hardly proceed at all without validity coefficients, or some reasonably adequate substitute therefor, the indices of which of necessity also require a criterion for their computation.

Having arrived at *n* "pretty good" items, each of desirable alternatives, we may hope, if the test is a time-limit test, to improve further the validity of each sub-test by utilizing scoring formulas, which, if we are to use them, "must *discount* errors" (i e, C must be negative in $S = R + C\ W$). This too presupposes a criterion because the weight of Errors, (E), of Wrongs, (W), i e., errors (E) plus omissions (O) relative to Rights, can be *ascertained* (as vs. judged arbitrarily) only by the aid of a criterion and the multiple regression equation The selection of tests for a battery presupposes that considerable numbers of sub-tests are tried out against the criterion as a sort of ultimate measuring rod, and that only "the few best" are chosen

To weight these chosen tests in a multiple regression equation then requires also a criterion The purpose here is to so weight the several sub-tests that a maximum validity of the composite or "scale score" results Even if the weights are arbitrarily chosen a criterion is needed in order to compute a validity coefficient of the thus arbitrarily weighted scale.

Finally, then, we may say that we require a criterion for the general purpose of maximizing or optimalizing all sorts of relationships in the problem of prediction; for determining the poor alternatives and choosing the better items, for determining the best scoring formula, for selecting the better sub-tests, for deciding the weights to be ascribed to each in a battery or "scale", for determining the validity coefficients of batteries or scales of tests weighted either arbitrarily or by multiple regression equations, for determining the "most causal" of all the predictive variables (inclusive of traits and environmental factors or variables and social relationships) and determining their relative weights; and for determining what basic form of relationship (including therein "higher forms of the multiple regression formula") underlies the best combination of all the "causal" variables For all such purposes a unitary criterion score for each person of our experimental population is indispensable.

Yet success is *not* unitary! A concrete consideration of cases will emphasize the point. Miss A is, let us say, + 1 σ in speed-of-typing, but is − 1 σ in accuracy-of-typing (that is, she "is speedy but inaccurate"). Miss B is − 1 σ in speed-of-typing but is + 1 σ in accuracy-of-typing (and popularly, "is slow but accurate"). If for each of these persons we add the scores, or average the scores, or (for these *particular two* persons) weight the scores *equally*, in all three circumstances they turn out to be 0 σ on the composite-of-speed-and-accuracy variable, or success-as-a-typist variable; that is to say, both are "mediocre." Yet by no reasonable stretch of the imagination are Miss A and Miss B equally "successful" typists The more true statement is to say that they have *a different kind*, or *a different type* or *a different pattern* of success. Most employers of stenographers will prefer for most purposes Miss B, because they hate errors, they hate the correction of mistakes and the complaints of customers and clients which inevitably result when errors are made. And besides, even if Miss B does less work, still very little of it has to be done over again. And doing work over takes much time—much more than enough to do it correctly in the first place—and costs money.[1]

Miss A and Miss B do have a different *profile* of typing success. This suggests that, possibly, if we could hold them to the same speed of production (that is, keep speed a constant) then their relative errors might measure their relative success, or conversely, if we could experimentally keep their errors a constant their resultant speed would yield us a unitary measure of their success; that is, that in either of the two alternatives we should have not two aspects of success but rather only one; and instead of having for each a two-traited profile we would have a unitary success score such as our ordinary regression and weighting equations demand.

A little thought will reveal the fact that this is not possible, psychologically, for even these *two* variables; for hold down the speed of Miss A and instead of becoming more accurate (as almost everyone presupposes will be the result) the converse actually may occur; while speed up Miss B and her accuracy in all probability will improve; at least the normal expectation is that, after a minimum of practice, such will occur. It all depends!

Also, speed and accuracy are not the only variables in success. Success is *not* a two-traited variable compounded only of speed and accuracy. If Miss A and Miss B are private secretaries their employer considers as aspects of their merit or worth other abilities as well: how well, for example, they can spell, to cover up his ineptness in this respect, the excellence of their grammar, so that they never say in a letter "have went" even though that may be what *he* dictated into the dictating machine; their ability and adroitness in answering the questions of visitors and in answering the telephone, and even their ability pleasantly and skillfully to entertain an influential customer until the boss's return In other words, even in simple jobs success is multi-dimensional Or, finally, we conclude that success on even a simple job *is* measured by the individual's profile on a multi-dimensional profile system composed of *m* criterion variables, where, generally speaking, *m* is greater than 1.

The upshot of all the argument to date, then, would seem to be that for an adequate treatment of the prognostic problem we should strive to predict the individual's success-profile, rather than to predict any unitary combination—*however skillfully combined*—of those criterion variables.

This prediction of a profile, as vs a unitary criterion score, may be done by the simple expedients of:

1 Administering enough tests of a wide enough variety to predict well (a relative term!) each of the several subportions of a criterion—the wages, the foreman's ratings of indispensability, the accuracy and the quantity of production over comparable times, and so on

2. Finding all the possible inter-correlations,

$$\frac{(m+n)\ \ (m+n+1)}{2}$$

in number, of the *m* criterion variables and the *n* test variables.

3. Employing each of the *m* sub-criterion variables in turn as a criterion and determining the appropriate regression equation for its prediction by the aid of the *n* several test variables in each case.

---

[1] An error, popularly, costs $X to produce it, $X to undo it (e.g., to erase it) and $X to do it over again, not to mention $X to locate it (inspection costs) Needless to say, the four X's are not all equal But this breakdown of the sources of the costs of errors gives some light on why errors are so abhorred by employers.

4. Making *m* different *Y* predictions for each examinee by means of the *m* thus determined regression equations; and subsequently plotting for him these *m* predicted criterion scores as a profile on a regular profile chart, the plotted profile thus being the predicted *profile of the person in question* [2]

Such a profile, in common with all profiles, reveals visually in relief in general three things.

1. The general trend, or tenor, of the scores.

2. The strong points and the weak points of the individual's several criterion "abilities"

3 The dispersion of the criterion "traits" about their central trend

To the extent to which the several sub-portions of a criterion are highly related (possess a common *G*, or mathematical factor) will the standing in the several traits be more or less identical and tend, consequently, to result in a series of ratings which are located in a horizontal line The normated entries (the encircled "ovals") of all columns of the plotted profile would be identical if all such predicted sub-criteria intercorrelated perfectly.

The mode of using such profiles—in selection, for example—is not generally agreed upon Consequently we are concerned more often than not with combining the several sub-criterion variables into a unitary criterion score, for in that case the simple concept of a "critical score" applies.

The success, then, of an individual, about which we prate so glibly, is a complex thing, and if it is to be made, artificially, *into a* unitary variable must be compounded of the weighted[3] sum of the several component parts, as one, simplest, conception of the matter. If we accept that definition the problems become three in number:

---

[2] *Caution.* Each such variable, Y_{1i}, will have a standard deviation, $\sigma_{Y_1} = R_{1.23\ldots m}\,\sigma_1$, and is therefore more constricted than is σ_{Y_1} Accordingly one will need to find by actual computation for each case a *Y* value and determine the distribution of that variable in order to find the decile ranges, the "norms," for example, for the profile sheet of column-1, dealing with aspect-1 of the criterion profile.

[3] It should be recalled that if we add scores, "using no multiplier," we really are using a gross-score weight of 1, which weights the several variables directly as the size of their standard deviations. Consequently if scores are combined they are weighted, and it is impossible, then, "not to weight" them.

1 To decide what variables are to be included in the criterion: i e., to decide what variables are to be weighted other than zero

2 To decide in what units—comparable measures—to record the individual's standing in the *m* several sub-variables, or component variables, of the criterion. Unless the several *sub-portions are recorded in* comparable scores, the multipliers of the scores will weight the variables other than intended

3. To determine how these are to be weighted, i.e., to determine by what specific weights we shall multiply in turn the several sub-criterion scores of John Jones before adding the several resultant products to obtain his final "weighted" criterion score. Let us consider each of the three problems in turn.

The Sub-Criterion Success Variables

Depending on the problem at *hand one will* have different success scores. Thus in a study of divorce, one would consider for inclusion in a criterion such variables as length of time that the marriage existed before the divorce, the happiness scores of the husband and wife at comparable times, or the sum or average of the two as an index or measure of the "happiness of the marriage," the number (and the quality) of the children in which the *marriage resulted before the divorce* (*if one is interested* in the eugenical aspects of divorce) and the like. The above statement suggests three corollaries.

1 That the *purpose* of the study, whether legal, moral, social, eugenical and the like, will in part determine, if not the individual variables, at least the *fields* or *realms* in which we may look for the variables to be included.

2. That there are always variables which *are on the border*land of "test" and "criteria"; ones which in one study are on one side of the equality sign (of the criterion-weighting equation) and in another are on the other Thus in a study of divorce in which the length of the marriage (stability or persistence of the marriage) is the focus of interest, the number of children logically is thought of as being a "test," the presump*tion (or hypothesis to be tested)* being that *where there are* children a marriage probably is more stable or lasts longer than where there are none. (There *could* be curvilinear relationship *between number-of-children* (*X*) *and the persistence in years* (*Y*) of the marriage) In another study where the center of interest is the social productiveness of the marriage, such as the number of children, the quality of the home, the wealth, happiness, inventions and personal adjustments secured, the number of children is certainly one of a number of alternative "products" of the marriages studied and quite clearly, then, is one *of a number of alternative criterion sub-variables.*

3 That there is no universal agreement as to what constitutes "success" in even one realm, not to speak of it as a generalized measure in all realms of life. Concretely, we do not agree even as to "why, or for what traits of success we pay a given man one wage and another another." Obviously, then, to say that "Mr. Jones is a successful man" means nothing definite *statistically*, *even if, by common agreement, it does mean* that the man in question has been successful in a financial way, probably that he has secured promotions, recognition and even acclaim; and possibly, though not surely, that he has lived a good and useful life—in other words, it suggests much but says nothing definite and positive Since there are no universally agreed upon definitions of success nor even concurrence as to what variables *to include, we must* be arbitrary in any case. *We may seek to mitigate the possible bad effects of our own judgment by pooling the judgments of other* (*competent*) *people as to what variables to include.*

For our purposes here it will be necessary to restrict the discussion to some one concrete realm. The one of greatest development to date is the vocational More criterion variables have been devised in this realm, probably, than in all others combined. Let us restrict our attention then to it.

In this realm, then, we may ask ourselves what are the *evidences* of *vocational success* Let us enumerate some of those which have been employed in such studies, while in order to save space we discuss each briefly at the point of mention

1 *Wages.* In free competition, where workers are all doing the same kind of work, and promotions "on the basis of merit" are promptly granted, wages evidently are a fair to good, if not excellent, aspect of a man's worth. A little reflection, however, will reveal that wages often are paid for quite other reasons than "merit displayed in free competition." One man is paid more than another of conceptually equal merit because he is a relative of the boss; because he has more dependents and "needs it"; because he asks for raises more persistently; because the boss thinks well of him, or simply "likes him", because the trade union to which he belongs has a minimum wage scale for persons of his "experience"; because he has been employed longer, etc., etc. Conversely two men of conceptually unequal merit may be paid the same wage because the company pays only a few different "day rates"; or because promotions come only at stated intervals and wages generally are most highly out-of-line immediately preceding a "promotion period", and the like.

Again wages often nowadays are paid according to a "wage formula" which "recognizes" many other factors besides basic "productivity." That is to say a wage, or a part of a wage, may be a bonus intended as much to *motivate* the person along *certain lines of* desired behavior as *to reward* him *for* "merit already displayed" Thus tardiness or absence from work may be penalized more heavily than the hours or minutes lost would warrant in the hope of inculcating "punctuality" and "dependability."

In general, *equal wages will reveal equal merit only when the attendant circumstances are equal,* where, for example, the *number of days or hours is equal; and,* in general, *where the "risks" are equal.* Where the wages are equal but the risk unequal, equal wage is not an adequate measure of the relative worth of the two men receiving it. Concretely, in the first of two sales territories it may be harder to sell goods than in another (of different "need," or demand, and socio-economic circumstance). Consequently equal commissions received in the two mentioned would indicate greater "sales ability" on the part of the man in the "more difficult" territory, The problem of saying precisely how much better is not necessarily solved by an arbitrarily devised correction.

Absences from work, if not allowed for, make the "pay envelope" an erroneous reflection of the individual's merit.[4]

Anything which arbitrarily reduces the standard deviation of the resultant sub-criterion scores has a deleterious, or attenuating effect on the resulting scores. Thus, in an effort to prevent "speed up" and consequent possible reduction of wages for doing a given amount of work, workers under bonus systems often stereotype the output; that is to say, all or all but a few do on any one day practically the same amount of work despite their ability in individual cases—possibly in a majority of cases—to produce much more Although the record of a day, or any number of days individually may show this, the *summated* record of a week or of a month may not be revelatory of this situation Under such conditions obviously the production is all but meaningless. The correlations with production of any truly good prognostic tests in this case will be attenuated, that is, greatly lowered Where uniform or practically uniform day wages are paid, soldiering on the job on the part of the more capable workers may have the same end result on any measures of quantity of production which may be collected.

Where workers work in pairs, aiding each other in simultaneous operations, the speed of the one will of necessity be the speed of the other, as in all "assembly line" production and any criterion, other than "errors," becomes largely meaningless The respective wages in this case are a measure of the "wage formulas" employed and not of individual differences of merit.

And one may be paid for "power to achieve" rather than for "achievement," because, for example, one *can* do hard jobs, or rare jobs or specialized jobs which another of the same labor force cannot. Even though this month both do the same work the first of two men may be paid more because he, alone of the two, *could* do the unusual job if one came up, that is, the employer is willing to, and does, pay for insurance that his firm "can handle all jobs that may come its way."

Often also there are notorious biases in wages. In Brazil it is said to be the custom for a worker to ask for, and confidently to expect, a raise with the advent of every baby in the family, and "ability" in this respect by no means is perfectly correlated with job-worth Again, women teachers, it is commonly believed, often receive far less wages for comparable services rendered than men, the theory, or justification, possibly being that "men support families" while "women work for pin money" and should be paid accordingly. And of course the old law of supply and demand still obtains! In such cases one may divide the investigation into two independent studies based on sex[5]

2. *Production.* Where all the workers of the experimental group are working at "exactly the same work," the quantity of work produced is a good measure of the merit of the individual man Production in general is what occupations are for. Yet current production, even on excellently planned "bonus work," may reflect only, or largely, irrelevant factors such as a man's temporary need for money, as for paying off a mortgage soon to come due, or reflect too largely his current condition as to sickness, fatigue, morale, or interest

And, *critical inquiry often reveals that the conditions of work are not equal,* that, for instance.

One salesman has odd jobs to do which subtract from the time in which he can give his full attention to his main job of selling.

One salesgirl, as *versus* another, is required to sell a higher proportion of obsolete stock, which consequently sells more slowly than the up-to-the-minute merchandise. Or she may have alternative tasks which affect her sales effectiveness, as, for example, to help keep stock in shape, to assist the buyer, and the like.

One man has a slightly higher gauge of metal so that his "poundage" reflects this fact, rather than his "assiduity"

Another man's machine, "just like this one," runs at a slightly higher rate of speed than his neighbor's so that his poundage, and consequently his pay envelope, reflect this fact.

The "units" of production, although assumed equal, may not in fact be equal. In punching 80-column Hollerith cards, for example, one girl may punch daily 1000 cards and a second the same number. If the first punches full cards (80 columns) while the second punches only 40 columns, the "column" measure of production reveals the first worker to be approximately twice "as good" as the second.

Even with the units (columns, in the above example) comparable the difficulty ("risk") of the tasks may be unequal so that 80,000 columns, or aggregate holes punched per day, of the one may not be equivalent to 80,000 columns of another. The one may be punching data requiring coding, which slows the task, or may be punching from a rather illegible data medium, while the other may have straight numerical punching from highly legible and exceedingly conveniently arranged data media. And even if both are numerical the one may have a data-medium where the entries are hard to locate (requiring many backward and forward movements of the eyes) while the second has a data-medium in which all the answers align in a column at the right so that punching them is all but an automatic job.

Time tends to relegate many factors such as those mentioned to the status of chance, or compensating errors. Accordingly, *the longer the time over which criteria are collected the less important,* generally but not necessarily, *are* such *"errors"*

In some cases, also, one may so arrange it that there are adequate controls of such matters, that, for example, half of the subjects work for one week on the one project while the other half work on the other, and the following week the groups interchange places, while some multiple of two weeks' wages is taken as the criterion variable.

It would appear, then, that production is likely to be a better criterion, if the attendant circumstances of its accumulation are under the control and supervision of the research worker; if, in a word, he collects the data himself, rather than taking uncritically what may be handed to him by a production officer or records clerk. Only a very few firms now collect routinely multiple, and reasonably comparable, aspects of the success of their workers on various variables.

3 *Quality of Work* The quality of work done by a "worker" is everywhere highly regarded as an important abstract virtue. In some cases, such as the performance of the stage, screen, opera, studio, prize-fight ring, and operating table, it is the paramount consideration Yet, like the former variables, it too has its difficulties

In bonus schemes bad quality of work frequently is discounted, and oftentimes more heavily perhaps than the "psychological" merits of the case would warrant. Such penalties often are "set high" as a means of discouraging their occurrence. Sometimes they are based on a quasi-rational principle as "the *estimated* time, often generous, which it would take to undo the incorrect or erroneous work and do it over again correctly or 'errorlessly.' "

Quality in most cases is based essentially on subjective human judgment. Limit gages sometimes give a highly objective status to quality measurements where the precise size of a product is an important desideratum in the finished product repetitiously produced Even here the man whose average error is .0002 usually receives no more wage, honor, merit, consideration, or even notice than he whose average error is 0005 when .001 is the allowable maximum deviation of the product from the ideal or perfect measurement One remedy for this is to devise measuring machines which reveal instantaneously, when a finished product is brought between the jaws of a measuring device and a constant pressure automatically is applied by means of a friction wheel, the amount of error on each piece, *and its sign.* This yields a *distribution* of the individual's actual errors the mean of which may be taken as his "inaccuracy." *Such a distribution customarily is not routinely collected by industry* It consequently can be had only by special inquiry—involving the specially designed machinery—introduced for that purpose

An objective "rating scale" may be constructed to measure quality. Thus by scaling cookies as to taste, and comparing the taste of each experimentally produced cookie with the taste of scaled samples one may ascribe numerical magnitudes to the taste of samples of cookies, and so assign the merit of the several cooks thereof, even though the scale be consumed in the process! Sewing, soldering, cookie-cutter, wire-splicing and lettering scales have been used and, when properly used,[6] yield objectivity and probably also corresponding validity, in their respective realms.

---

[4] One must still heat the factory even if a certain employee is absent, justifying considering somewhere—but not necessarily in wages—this criterion component of "dependability"

[5] An alternative is to employ sex (the categories of which are arbitrarily quantified) as a test variable. Still another is to make some arbitrary correction for sex, thus to add to all females' wages a constant, the difference found by subtracting from the mean male wage the mean female wage. There is no unanimity of opinion as to what is best to do in such situations. This field of research procedure needs a careful scrutiny We need criteria for deciding such research dilemmas.

One may sample the products at periodic intervals, by noting, for example, the errors made in successive intervals, collected at any time in hourly intervals by the inspector on his rounds in successive, frequent, but unanticipated intervals throughout the work day. This has been called the timed sampling technique Errors in this may arise if the worker strives, consciously or otherwise, to do a better quality of work for a limited time when he suspects an inspection is "about due."

The best measures of quality result from those situations, in repetitive production, where all errors above an allowable minimum are automatically detected by objective inspections (e.g., by the electric eye). In this case the poundage of discarded work is a significant measure of the quality of work done.

Breakage and spoilage of raw materials or of partly finished products to which the worker is fitting additional parts are important aspects of a man's worth and so often are elements of a criterion. In assembly or bench work often these constitute the core of the criterion To take account of different speeds of assembly of different workers one may take as his index of breakage or spoilage the percentage (or proportion) which the wasted material is of the total material processed.

Clerks who otherwise are ideal employees may fail miserably in their employer's eyes *if they are highly susceptible to making* many clerical errors, illegibilities of writing, or errors in computation, particularly the undetected ones which cost the employer money or require him to adjust matters with irate customers.

4. *The Rate* (*or, alternatively, the amount*) *of Acquisition of New Skills.* In aptitude tests, generally, the length of time

---

[6] It is often overlooked that if it takes *m* judges to determine an objective scale it may also take almost as many judges reliably to ascribe scores to individual products by its aid.

taken by persons, equally unskilled at the beginning, to achieve equal competence is a rather valid criterion measure. This presupposes "equal opportunity to learn" It stresses the overhead aspects of employee worth In many jobs the training is costly. In war-time, for example, delays in production—due to training time, or any other cause—are scanned more closely than usual Other things equal, that employee who costs least to obtain, to train and to maintain, is the most valuable. And if promotion is contemplated, the fast learners are those most worth retaining. There is at least a fair presumption that they will be rapid learners on the subsequent job employing a good core of the same kinds of skills.

This criterion may be applied either to apprentices in a formal training school or to workers in training on the job itself. In fact, as will be shown below, it is sometimes possible to so arrange it that the job itself becomes a prognostic test with generally good all-round results.

5. *Supervisor's Judgments.* These may be of "over-all proficiency" and may be employed in lieu of a more satisfactory objective criterion; or they may be ratings, say, of the several specialized ends or objectives of the work done, of which it is generally claimed, or claimed for the present, in lieu of objective scales thereof, that only the foreman or other supervisory officer is an adequate judge.

These often are notoriously unreliable in the technical, as vs. the popular, sense The judgment of a foreman in an eyelet department of a certain brass factory agreed with those of his assistant as to the over-all merit of his men to the extent of only 60, even though aided by a formal card-selection-into-piles *method of rating. He agreed only slightly better with himself* on a subsequent occasion; that is to say the self-correlation was but little better. The antidote for such, of course, is repeated ratings, monthly or bi-weekly or even weekly ratings over a considerable period of time, with all the resultant ratings aggregated into a composite, summated or averaged, score By this expedient the composite rating may be made to approach as near *1.00 in reliability* as we wish while *its validity* may be expected to obey the Brown-Spearman prophecy formula; that

is to say, may be expected also to increase somewhat but to approach rapidly, with increase in the number of ratings, a fixed *ceiling which generally is far short of 1.00 The crucial questions* here are: (a) just what "traits" shall be so rated, and under what circumstances? and (b) how many ratings should be secured, and from whom?

The rating of the products of artists' work, pictures, etchings, and the like, are notoriously unreliable, intercorrelating no higher, perhaps, than .30. Even the reliability of the much-vaunted medical skill and judgment in diagnosis is probably no higher than 50. In the latter case Brown's formula tells us that to secure a composite judgment with a reliability of .95—such as any group intelligence test, in its field, will yield on a 2-hour examination or thereabouts—not to mention *validity,* would require the *independent* judgments of an assemblage of some nineteen equally competent doctors It follows that (a) often one cannot assemble enough "judges" who know well enough the subjects of the investigation to rate them accurately, and (b) that one often cannot, in many fields, obtain judgments or ratings "worth their salt."

The reports of a training supervisor may be of more worth than of a regular line officer Thus the worth of an assistant foreman, for purposes of rating job-worth of employees, may be greater than that of his superior, the foreman-in-charge, particularly if the former supervises the men in their day-to-day work in the shop, while the latter concerns himself largely, or at least more than the former, with the general planning of the work and the paper-work of the department. In fact the latter, under this circumstance, may be all too much influenced in his ratings by his knowledge of the wages which they receive and all too little by their more pertinent behaviors under varying shop conditions The size of the pay envelopes or the exaltedness of the titles of the supervisors does not necessarily correlate highly, or even positively, with their "ability to judge." The correlations *could* be negative!

If the foreman believes that "experience makes the man," his ratings or *rankings* will be substantially a ranking of the men by age, or by experience And age, or experience, in this

case will correlate more highly with his ratings than any other predictive variable Since age, or experience, in most occupations does not correlate highly with "competence" it follows that this attitude leads to spurious or attenuated ratings the badness of which will not be readily apparent.

It is notoriously true that supervisors of teachers are not able to rate reliably—not to mention validly—the teachers under their direction on even a fairly *objective* trait,[7] such as the teacher's "ability to gain and hold the attention of the class" Consequently we must believe that any generalized objective such as the "teachers' teaching effectiveness" cannot be reliably rated, even with the aid of formal rating scales, without many repeated visits for observation of the actual teaching. The ratings actually received often are a composite of the rumors about a teacher seasoned only with a dash of real knowledge thereof.

The *situation in school or industry in* which *more* than two—or even two—persons know intimately all those under their direction is rare indeed Two ratings by different persons are probably more valid than the same number, two, of repeated ratings by one judge, but if the former cannot be obtained the latter is about the only alternative

One thought here, however, where the number of judges of *necessity is inadequate, is to have fellow-workers, fellow*-apprentices, fellow-teachers, or fellow-pupils, say, rate one another. The extra number of raters may either partly or fully compensate for any assumed "lack of experience in (or competence of) judgment." Such "judges" indeed are able, for example, to note certain aspects of competence not so readily detected by the supervisory officer. Workmen often know, for example, when the shoddy work of a fellow-worker is being "buried" by an overdose of a beneficently-hiding coat of paint, like the doctor's *mistakes* by the casket!

One might even weight more heavily the judgments rendered by the adjudged better workers than those of the poorer;[8]

---

[7] Bowman, Earle. *A Plan for Evaluating Teaching in Terms of Pupil Activities.* Unpublished Dissertation. Ohio State University Library, 1928. Pp. 185.

[8] This presupposes a positive and "high" correlation between ability to do and ability to judge. Several studies have in fact shown that this is the case; that,

but if "everybody judges everybody"—under conditions where nearly everybody knows everybody well—it is a fair guess that the additional returns from such weighting generally may not pay for the additional time and trouble necessary

6. *Knowledge* Achievement tests and trade tests of knowledge possessed may be employed as a criterion sub-variable There is more truth than fiction in the standing joke that the eminent surgeon's bill of $200 for removing your appendix really is $25 for "removing your appendix" and $175 for his "knowledge of what to do and what not to do" Popular opinion has it that there is an almost zero correlation between knowledge and skill Actually in most trades the two correlate quite highly; so highly in fact that oral trade tests (measuring knowledge primarily) are excellent tests for picking men in order of their ability "to perform" on the job.

7. *Job Tenure.* This is comparable if the employees, say, all entered the firm at the same time and so have a good chance of being subjected to the same (historical) factors and conditions. This variable is only slightly related to "competence." It probably is highly related to that aspect of job success known as "job satisfaction." It would not be comparable if business cycles, wars, or other "disturbances" did not "cover" all employees' records equally.

Following general reasoning that that profile is most stable (statistically reliable)[8] and most representative of the person, other things equal, which is most unique (in the factor sense) it would seem that it could be concluded that one should aim in predicting profiles to include as many unique variables as possible, while if one is combining them into specific sub-criterion variables, such as "accuracy," then the variables included should be as "loaded" with the sub-criterion "factor" (here "accuracy") as possible. In the second case, the accuracy "tests" should be highly intercorrelated rather than the converse.

8 *Supervisory and Leadership Ability* The normal progress of a man in industry—and particularly so in war times—is to be given supervisory responsibility as fast as he demonstrates an ability to assume it. Where this is the case the *supervisory status* of a man is an important criterion element. The foreman is distinguishable from the workman often, or even usually, not so much by "superior ability to do" as by superior knowledge, better judgment, greater ability to solve difficult practical and theoretical problems, ability to lay out work, to size-up repair jobs, to criticize faulty work, and to induce men cheerfully and uncomplainingly to work industriously

The number of workers supervised is probably of far less moment than the adequacy, or quality, of the leadership rendered This conceivably might be measured by some form of check list.

9. *Incidental Factors.*

(a) Amount of supervision required.

(b) Attitude toward supervision

(c) Job satisfaction.

(d) Adjustment to the work and to fellow-workers.

(e) "Influence" on fellow-workers (e.g., morale-building influence).

(f) Skills inventory, for example a measure of the "correctness of motions," the extent of possession of such a set of motions that present efficiency is likely not only to persist but also to increase (improve with practice). Does this auto driver have the right "driving habits" as measured by a check list? Does this salesgirl have the right sales tactics, as ascertained by a professional shopper? Does this professional swimmer have the "logically correct swimming movements"?

Viteles points out that success is not speed plus accuracy plus job tenure, but that it is an integrated whole. The difficulty of explaining an *n*-dimensional whole in quantitative terms requires as the only alternative to combining the several scores, as by weighting, for example, that one treat the success as a profile[10] as above indicated.

Bingham and Freyd have pointed out that *objective criteria* are more likely to correlate with objective tests, while *ratings* are more likely to correlate highly with tests of personality. If so, ratings, often hitherto employed as criteria, no doubt have perpetuated some tests which ought to have died a natural death long ago.

A criterion that is objective and uninfluenced by human judgment is usually "a more predictable measure" of job proficiency than a rating based entirely upon supervisors' opinions of the workers' performance.

A few general rules applying more or less to all such criterion elements may not be out of place·

It is important to make provision for the orderly and systematic collection of the criterion first, that is, ahead of administering the tests; otherwise the subjects may "get away" from one (by the route of graduation or drop-outs of students, the transfer of workers, the shipment overseas of soldiers, and the like), whereupon one will find that he has a fine bunch of expensively collected and scored tests papers, but no criterion to tell him whether they are worth anything or not.

The test scores cannot of themselves tell one whether they are of any worth. One of course must have a wide dispersion of test scores, but that of course is a fairly common characteristic of worthless tests. The test must be fairly difficult, ideally of such a difficulty that the average score is about half the possible maximum score, but that also is a characteristic of many worthless tests It must have some reliability, but, in the experimental stages and other things equal, the reliability had better be low rather than high to the end that its validity as well as its reliability may be greatly improved by lengthening it. Among existing tests generally—and possibly even among tests of a particular type as well—there is no correlation of note between reliability and validity Accordingly high reliability in itself warrants nothing as to the value of a test. And finally, even if a test has a large dispersion, the right difficulty and a good to high reliability—all three—even that *pattern* of suppositions merits nothing as to the value of a test. Validity may be established only by a validity coefficient, in turn determinable only by the aid of a criterion.

Much time and care should be expended on the collection of adequate variables. The final test scale or prognostic battery probably will be no more valid than the yardstick by which it was constructed. And at best it can be no more valid than the square root of its reliability coefficient. This conclusion is achieved by manipulation of the formula for the attenuation coefficient.

Possibly as much time should be spent in devising the criterion as in constructing and perfecting the tests. This important part of a research seldom receives half the time or attention it requires or deserves If the criterion is slighted the time spent on the tests is, by so much, largely wasted.

All criterion variables should be collected over comparable times The measurement of *improvement,* for example, in general should start at comparable periods in the several persons' individual learning or growth curves

Criterion variables which accrue in time should represent the same amount of time sampling. It is better, for example, to include in a study all freshmen who persist to the end of the year, taking as a criterion score for each his end-of-the-freshman year scholarship (to one definite point in time), than to include also therein the averaged final success of those who dropped out with only one quarter (one-third of a year) of college work and also the final success of those additional ones who dropped out with only two quarters (two-thirds of a year) of college work and so on The three populations mentioned have over-all marks which have vastly different reliability coefficients, say 70, .824, and 875 (Brown's reliability formula), respectively, and considerably different validity coefficients. This is partly in the interest of rendering comparable the reports, for example, the validity coefficients, regression equ

---

indeed, this is a point which education thus far has but little recognized or exploited, namely that at a certain stage in the progress of learning one may possibly secure more progress by having the learners rate the products of their fellow trainees rather than do another construction project. See Smith, R. E. and Toops, Herbert A. "An Experiment in Self-Rating Shop Products." *Industrial Arts Magazine*, XV (1926), 225-230.

Rating, it may be assumed, develops the critical faculties of the workman; and after a modicum of skill at performance has been obtained the ability to detect the faults of one's production—as a result of the training in observing the faults of the work of others—obviously may be the royal path to their early and subsequent elimination in one's own work.

[8] Edgerton, H. A., Bordin, E., and Molish, H. "Some Statistical Aspects of Profile Records." *Journal of Educational Psychology*, XXXIII (1941), 185-196.

[10] Even those profiles which may be adequately denoted by code numbers (from [illegible]) can scarcely be apprehended in entirety.

tions, and the like. If one employs the maximally heterogeneous group, the validity found by another who uses the same tests, otherwise comparable, would be exactly comparable only if he had the same *proportions* of the three sub-populations respectively

If data are missing in only a small portion of the cases and in only a few of the criterion variables these may be and perhaps ordinarily should be supplied before beginning to combine the data Toops[11] has developed a formula for this The formula assumes that for any missing score in a given variable one should supply the weighted average standard score of the person in question on those variables in which the data are available. If the variables are highly intercorrelated, and few scores are missing, the assumption is justified. If a small per cent, preferably not over one per cent, of the persons cannot be rated, or the ratings (rating slips, for example) are lost, and there is reason to believe that these are a *random* selection of the total, then the data-present cases may all be reduced to centile ranks on the basis of the available N's, while the missing-data cases may be assigned the rank 50.[12] To facilitate such "transmutation" on a large scale one may employ Hull's Tables.[13]

One may in some cases devise a mechanical device to count the success of workers automatically, possibly even unknownst to them. Thus a Veeder Counter may be so attached to a typewriter or a Hollerith punch that it counts every key-stroke produced. If the one machine is employed daily and continuously by one operator, and one operator only, the total strokes of any one day, week, or month, as the case may be, is equal to the close-of-business reading less the reading-at-the-beginning-of-the-time-interval Over long periods of time even errors, such as the use of the machine occasionally by an executive after hours to type a brief letter and the like, tend to iron out If such machines are taken over by substitutes when the employee in question is ill or absent this may introduce appreciable

---

[11] Toops, Herbert A. "The Selection of Graduate Assistants," *Personnel Journal*, VI (1928), 470–471.

[12] Theoretically half by chance should be assigned the rank 50 and half the rank 51

[13] Hull, Clark. *Aptitude Testing*, Appendix 1 Yonkers World Book Co., 1928. Pp. 491–492

errors into the resulting scores, but less than would be produced by absence and non-use of the machine by anyone if the days not worked are not allowed for

Most systems of allowances for errors are not particularly psychological By means of addends the patterns of errors (or of correct responses) may be recorded[14] and later may be subjected to alternative modes of analysis. To all intents and purposes the addend code number *is* the original performance.

Rate-setters and time-setters have many ridiculous customs and "allowances" as well as wise and wiser ones Their "standard time," may be the shortest of rough guesses rather than the result of a painstaking analysis They often are overly influenced by a "star (99 percentile) performance" rather than a statistically determined average performance.

In criterion-building we often imply the equality of numerous "things" which in fact never are equal·

1 That the motivation of the subjects is equal. This is a particularly important consideration.

2. That the "risk" is equal We here are confronted with such questions as. Do people ride cabs equally frequently in this cab driver's "district" as in others? The distance traveled or the fares collected may not be an adequate measure if that is not the case If the houses are farther apart (as in the newly-built-up sections of town) the potential riders per mile are fewer but the actual riders per mile may be more.

Is work available at all times so that this machine tender's wages are free from error by reason of his pay envelope not reflecting, or reflecting inadequately, "lost time"?

3 That the experience, or practice, or education, is equal.

4 That the human factor, as vs. machine factors, is the one which produces the basic variation in criterion scores of the various individuals of the criterion group

5. That the persons employed for a criterion may allowably be combined in a society, rather than properly only be divided into two or more. We have mentioned sex above. The same can be said for age, and in the case of machine-tenders possibly the age of their machines.

---

[14] Toops, Herbert A. "Code Numbers as a Means of Scoring Group-Administered Performance Test Products." *Journal of Applied Psychology*, XXVI (1942), 136–150.

6. That the work environment is equal Ventilation, noise, lighting and other conditions all have differential effects on individual workers which in some instances amount to biased errors, or worse. Racial or religious antagonisms sometimes divide any industrial department into "good" and "poor" producers

7 That the progress of work is not impeded by hindering factors, such as unavailability of raw materials, transportation, power or supplies over which the employee has no control

Comparable Scores

Before the several portions of a criterion to be combined are weighted, each person's X-score in each variable must be reduced to comparable scores. For this purpose comparable scores may be defined as any scores that have equal variability. Since "success" at best is an arbitrary variable, one need not involve in the definition the requirement of equal means in the several variables. If this is the definition adopted, then rankings (based in common on all N persons in each different criterion variable, with "tied scores" outlawed) are comparable scores; and so are gross standard scores or Z's $\left(Z_1 = \frac{X_1}{\sigma_1}\right)$, and also A.D. units, median deviation units, and the like.

Because of their highly desirable algebraic properties, the ability to predict exactly standard errors of estimate, for example, if they are used, practice restricts comparable scores to standard scores or some variations thereof, such as Z's, or T-scores, and the like, all of which in common have equal standard deviations. Standard scores, $z_1 = \frac{x_1}{\sigma_1}$, are most frequently employed and have a property, sometimes valuable, that all the criterion variables, automatically by its use, are made to have equal means as well as equal standard deviations.

The Arbitrary Weighting Formula

The arbitrary weighting formula for combining the m several sub-portions of a criterion into a criterion variable, or the

several sub-criterion variables into a unitary criterion score, accordingly is:

$$Y_c = \frac{\beta_1 X_1}{\sigma_1} + \frac{\beta_2 X_2}{\sigma_2} + \dots + \frac{\beta_m X_m}{\sigma_m} \tag{1}$$

where $X_1, X_2, \ldots, X_m$ represent the m several sub-portions to be summated, and $\beta_1, \beta_2, \ldots, \beta_m$ are *arbitrary* weights, not precluding "logical" weights, to be accorded the m several portions, or variables, respectively

If one is going to predict a profile instead of a unitary criterion score, one will naturally desire to secure as adequate an "accuracy score" or "speed score," for example—one of the basic variables of the profile—as possible. It follows that possibly one will procure several different accuracy scores for each individual, e.g., his September accuracy score, his October accuracy score, and his December accuracy score. In this event, then, we shall need to combine two or more scores (in the above case three) for each "variable" (such as speed, accuracy, bonus earnings, etc.) available. Consequently in this case we still will have the problem of combining scores, and the above formula with corresponding adaptation still is appropriate.

Thus we cannot escape the "weighting" dilemma. Even if we have only one criterion score, we still would weight the variable with $\beta_1 = 1$, in the bids system below.

Reversing Signs of Sub-Variables

If one is combining "accuracy" scores and "error" scores one may reverse the signs of the "error" scores so that they will combine properly with the former This can be done either by giving a negative sign to the β's or to the error scores, in the bids system below. One of the best systems is to transmute all the original scores, X_1's, to transmuted scores on such a variable by the formula,

$$X_1 = K_1 - \beta_1 X_1, \tag{2}$$

where β_1 is the bids-importance of the trait, a positive magnitude, and K_1 is taken of such a magnitude that all resulting X_1'-scores are *positive*. It follows that the resulting X_1''s will all be positive but that the less meritorious (e.g., more erroproducing persons) will have the smallest-sized scores and *vice versa.*

The Bids System of Arbitrary Weights for the Portions of a Criterion

The several arbitrary weights $\beta_1, \beta_2, \beta_3, \ldots, \beta_m$ of formula (1) most conveniently can be obtained by the bids system proposed by T. L Kelley for weighting factors-of-merit in rating the S.A.T.C. in World War I. In essence it involves the following elements:

a *N* judges, at least above a minimum of competence, ascribe to each of *m* criterion variables bids-of-relative-importance, in respect to their judged importance, under the controlling principle that the sum of the bids allotted shall exactly equal a certain predetermined quantity, say 100 The situation is thus rendered as close as may be to that of, "all other things equal, what per cent of the total importance (variance) can be ascribed to a deviation of 1x, in this variable of concern " It will be noted that this is one of the central points in our notions of variance. Consequently it is to be presumed that statisticians who have dealt with this specific problem, after an adequate on-the-job study of some weeks or months, may be more capable of doing what is statistically demanded than the run-of-the-mine foreman or supervisor who is about the only available other source of "expert judges"

b. After the judges have rendered *independently* their verdicts of bids on the several traits, preferably in integral numbers adding to 100, one may weight all the judgments of a given judge, if desired, according to his assumed *competence-at-judging* Thus one could weight the judgments as the square root, or cube root, say, of the judge's years of experience in *supervision* up to some maximum—say 20 years. If several expert judges are available the simpler procedure is to assume that all are equally competent, and accordingly by merely adding the scores one in a sense weights them equally.[15] Thus in every

---

[15] Since they all summate to 100, thus it would seem tends in part at least to weight the several judges equally Of course the weights may add to a predetermined sum and yet the standard deviations of the bids may vary considerably, particularly in the case of erratic judges. Where scores are added that judge who has the largest standard deviation, of course, receives most weight.

case we obtain a weighted (summated) criterion score for each individual

c. A compromise set of bids is arrived at, possibly the rounded averages of the several judges' bids on a given trait. No harm is done, and some gain in interpretation is secured, if these are so rounded that their sum also is 100

d The compromise bids thus secured are the values of $\beta_1, \beta_2, \beta_3, \ldots, \beta_m$ of formula (1).

The judges should pay no attention to the *signs* of the β's, as in the case of "accuracy" and "errors" above, but should judge the traits as-if-they-were-of-absolute-sign.

Points to be noted in the ascription of the bids[16] are·

(1) Variables repeated in other variables (having a high correlation with other variables) should receive a low weight.

(2) Variables which are subject to error, other things being equal, should receive lower weights than those not so subject to error

(3) Variables representing an adequate sampling over a long period of time should receive a high weight, other things being equal, relative to a variable which is a sampling of only a short performance of a similar trait.

(4) The bids should be made up independently, i.e., without the judges conferring during the original distribution of bids

(5) After each of two or more judges independently has distributed 100 bids to the *m* traits, as much revision, before comparing results, as desired should be allowed.

(6) After the bids are thus secured, a set of compromise bids should be made up. These likewise should add up to 100. In this construction of the compromise bid, the above first three principles, and all other pertinent ones, should be invoked in the discussion, to insure that a fair set of bids result. It may be found, for instance, that two ideals clash; the use of assistantships, etc., as rewards for the encouragement of good scholarship, and their use to obtain cheap labor for departmental routine duties, for example. It may be found necessary to make some distinction between these two; even to the extent of establishing two sets of weights, and evaluating each candidate's merits from the two points of view, successively It would seem to be good departmental policy, in that event, to evaluate for both appointments every candidate applying for either of the two types of appointments, irrespective of the fact that the candidate applied for only the one appointment. Otherwise, from the department's point of view a capable assistant, for example, may be lost if there are not enough scholarships and fellowships "to go round", while from the student's point of view, often he will find the alternative position offered him to be very acceptable.

---

[16] Toops, Herbert A. "The Selection of Graduate Assistants." *The Personnel Journal*, VI (1928), 457–472.

PERSONALITY AND INTEREST FACTORS IN DENTAL SCHOOL SUCCESS

CLAUDE EDWARD THOMPSON
Northwestern University

THE purpose of this study was to determine whether or not certain criteria of success in Dental School are significantly related to scores on personality and interest scales. Previous research by the writer (15) on the relationships between motor and mechanical abilities and success in Dental School required that he interview personally over one hundred practicing dentists and the faculty of a College of Dentistry. These men were almost unanimous in claiming that personality and interest factors are of as much importance as aptitudes in determining success in dentistry.

Published reports of research and reviews of the status of selection and counseling techniques in dental schools (1, 2, 4, 5, 6, 7, 8, 12, 13, 14) indicate also that measures of personality and interest would be of value in selecting and counseling. The *Strong Vocational Interest Blank for Men* is being used as a predictor item for success in dentistry at the University of Maryland. However, no published reports are yet available to reveal relationships between scores on the interest scale and criteria of success in this school.

Test Used

During 1942–43, three tests were administered to students in the College of Dentistry at Northwestern University[1] These tests were

(1) *Preference Record—Form AS*, by G. Frederic Kuder (3, 9, 10, 17).

---

[1] The writer is indebted to Harold A Grever, Director of Admissions of the College of Dentistry, Northwestern University, for assistance in the test administration

(2) *California Test of Personality—Adult Series*, devised by Ernest W Tiegs, Welles W Clark, and Louis P. Thorpe (16).

(3) *MacQuarrie Test for Mechanical Ability*, by T W. MacQuarrie (11)

The *MacQuarrie Test for Mechanical Ability* and the *California Test of Personality* were administered to 158 freshmen. The *MacQuarrie*, the *California Test of Personality*, and the *Kuder Preference Record* were administered to 66 seniors

Criteria

Previous studies (7, 8) at the College of Dentistry, Northwestern University, indicated that the *MacQuarrie* was a usable test for predicting freshman year average grade. Following a suggestion in those studies, an attempt was made in the present study to refine criteria for test evaluation by separating theoretical and technique grades from practicum grades This was done by obtaining cumulative points earned in technique and theory and cumulative points earned on product or work done. The scores were arranged in ascending order and divided into deciles It was then possible to score anywhere from 1 to 10 in both kinds of grades.

Results

Table 1 presents the correlations obtained for the 158 freshmen between theory and technique and practicum criterion scores and scores on the two tests The *California Test of Personality* gives Self-Adjustment, Social-Adjustment, and Total Adjustment scores These were correlated separately. It should be pointed out that the *California Personality Test* (16) has not been validated against objective outside criteria The 180 items were evaluated in the following manner·

TABLE 1

Correlations of Test and Criterion Scores for 158 Freshmen

	Theory and Technique	Practicum
MacQuarrie Total Score	.05	.11
Self-Adjustment Total	-.04	-.09
Social-Adjustment Total	.08	.20*
Total Adjustment Score	12	.20*

* Correlations approaching statistical significance are indicated by asterisks in all tables.

(1) Judgments of teachers, principals, test experts, personnel directors, and employers as to whether or not each item was an indicator of adjustment and employability

(2) The reactions of employed adults as to whether or not they judged each item to be an essential characteristic of a successful employee.

(3) The extent to which the results of the test agreed with the known characteristics of particular adults.

(4) The extent to which each item was consistent with the score on the test as a whole (bi-serial r)

Due to the questionable reliabilities of the six components in each of the categories Self-Adjustment and Social-Adjustment, it was decided to use only the over-all scores for purposes of group comparison. The split-halves reliabilities of these totals are:

Sec. 1	Self-Adjustment	.888
Sec 2.	Social-Adjustment	.898
	Total Adjustment	918

Only two of the correlations approach statistical significance For the 158 freshmen there is no relationship between total score on the *MacQuarrie* and standings in the criteria There is positive but low correlation between Practicum criterion scores and Social-Adjustment and Total Adjustment scores on the *California Test of Personality*.

Table 2 presents the correlations between criterion scores and test scores for 66 seniors. Only the three of nine components of the *Kuder Preference Record* on which seniors averaged definitely above the norms for this scale were correlated with the criterion scores.

For both freshmen and seniors (Tables 1 and 2) the *MacQuarrie* total scores have little or no relationship to either Theoretical and Technique or Practicum criterion scores. It is possible that the first three sub-tests of the *MacQuarrie* measure motor skills and the last four sub-tests measure mechanical ability Using total scores might, therefore, obscure relationships of sub-tests to criterion scores Total scores were computed separately for the Tracing, Dotting, and Tapping tests and for the Copying, Location, Blocks, and Pursuit tests These sub-total scores were then correlated with the criteria These correlations are presented in Tables 3 and 4.

TABLE 2

Correlations of Test and Criterion Scores for 66 Seniors

	Theory and Technique	Practicum
MacQuarrie Total Score	17	13
Self-Adjustment Total	.23*	.26*
Social-Adjustment Total	.20*	.22*
Total Adjustment Score	.22*	.26*
Mechanical	-10	-.06
Scientific	.28*	11
Social Service	-.01	.24*

TABLE 3

Correlations of MacQuarrie Sub-Total Scores with Criteria for 158 Freshmen

	Theory and Technique	Practicum
Tracing, Tapping, Dotting	-17	.22*
Copying, Location, Blocks, Pursuit	.16	-.23*

It can be seen that consistently, for both freshmen and seniors, Tracing, Tapping, and Dotting scores correlate negatively with Theory and Technique and positively with Practicum scores, and Copying, Location, Blocks, and Pursuit correlate positively with Theory and Technique and negatively with Practicum scores. Five of these eight correlations approach statistical significance. If the first three sub-tests measure motor dexterity and the last four sub-tests measure perceptual or visualizing abilities, Theory and Technique scores could be expected to correlate with the last four sub-tests and Practicum scores could be expected to correlate with the first three subtests. This expectation receives support from the positive and negative sign directions of the obtained correlations, though such breakdown does not provide correlations of sufficient magnitude for use in individual prediction. All correlations (Table 2) between scores for seniors in Self-Adjustment, Social-Adjustment, and Total Adjustment scores on the *California Test of Personality* and Theory and Technique and Practicum criterion scores are positive but low. The correlations between Practicum criterion scores and Self-Adjustment, Social-Adjustment, and Total Adjustment scores approach statistical significance The correlations between Self-Adjustment, Social-Adjustment, and Total Adjustment scores and Theory and Technique approach statistical significance. These findings indicate rather consistent relationships between what is measured by the test and standings in the criteria. However, these correlations would not be useful for individual predictions.

TABLE 4

Correlations of MacQuarrie Sub-Total Scores with Criteria for 66 Seniors

	Theory and Technique	Practicum
Tracing, Tapping, Dotting	-19	.32*
Copying, Location, Blocks, Pursuit	.22*	-.27*

TABLE 5

Average Percentile Standings of 66 Seniors in Components of the Kuder Preference Record

Components	Average Percentile Obtained
Mechanical	91
Computational	50
Scientific	93
Persuasive	30
Artistic	45
Literary	21
Musical	25
Social Service	67
Clerical	8

Table 5 presents the average percentile standing of the 66 seniors in the components of the *Kuder Preference Record*.

The average senior scores above average in Mechanical (91 percentile), Scientific (93 percentile), and Social Service (67 percentile) on the *Kuder Preference Record*. Mechanical interest scores do not correlate with either Theory and Technique or Practicum criterion scores (Table 2). Scientific interest scores correlate positively and significantly with Theory and Technique criterion scores but not with Practicum, and Social Service interest scores correlate positively and significantly with

Practicum criterion scores but not with Theory and Technique. These findings indicate that interest patterns are related to the marks earned in Dental School, but the relationships appear to be more specific than the relationships between personality measures and marks earned

The failure of the scores in the three components of the *Kuder Preference Record* to correlate more highly with criterion scores could have been due to the narrow spread of the scores (see average percentiles, Table 5), particularly in Mechanical and Scientific interest scores. It is well known that the coefficient of correlation is affected by the variability of scores in the group tested To test this idea adequately it will be necessary to administer the *Preference Record* to a group of freshmen and a group of seniors in a follow-up study

The deviation of an individual's interest scores from pattern (Table 5) may also be an index of the extent to which the scale places those freshmen and seniors standing low in Mechanical, Scientific, and Social Service interest scores low in criterion scores and places those scoring high in these same components high in criterion scores.

To determine whether or not there were statistically significant differences between the mean scores of freshmen and seniors on the *MacQuarrie* and the *California Test of Personality*, critical ratios were obtained These critical ratios indicate that there is no significant group difference between freshmen and seniors on the *MacQuarrie*. There is a significant difference in favor of the seniors in Self-Adjustment scores on the *California Test of Personality*. There are 94/100 chances that a difference in favor of the seniors in Social-Adjustment scores is real. There are 99/100 chances that a difference in favor of the seniors in Total Adjustment scores is real

Summary

(1) Seniors in dentistry score above average in Mechanical (91 percentile), Scientific (93 percentile), and Social Service (67 percentile) interest scores on the *Kuder Preference Record.* Mechanical interest scores do not correlate with either Theory and Technique or Practicum criterion scores. Scientific interest scores correlate positively with Theory and Technique criterion scores and Social Service interest scores correlate positively with Practicum criterion scores.

(2) Statistically significant differences in favor of the dentistry seniors over dentistry freshmen in Self-Adjustment, Social-Adjustment, and Total Adjustment scores on the *California Test of Personality* were obtained Whether or not these differences are due to selection, age, training, or all three is not known

(3) The *California Test of Personality* gave statistically significant positive correlations between Practicum criterion scores and Social-Adjustment and Total Adjustment scores for freshmen and between Practicum criterion scores and Self-Adjustment, Social-Adjustment, and Total Adjustment scores for seniors Statistically significant positive correlations between Theory and Technique criterion scores and Self-Adjustment, Social-Adjustment and Total Adjustment scores were obtained for seniors.

(4) The results of this investigation indicate that correlating *MacQuarrie* total scores with criteria may obscure relationships of sub-test scores to criteria

Personality and interest scale scores show some relationship in this study to criteria of success in Dental School, but the correlations are not of sufficient magnitude to be useful in individual prediction when selecting applicants for admission to the College of Dentistry.

REFERENCES

1 Bellows, Roger M. "The Status of Selection and Counseling Techniques for Dental Students." *Journal of Consulting Psychology*, IV (1940), 10–15.

2. Bronner, F. J "Preliminary Report on Aptitude Tests Correlated with Dental Subjects." *Proceedings of the American Association of Dental Schools*, XII (1935), 211–213

3 Crawford, A. B Review of the *Preference Record* in the *1940 Mental Measurement Year Book*, edited by Oscar K Buros. Highland Park, New Jersey, *The Mental Measurement Yearbook*, 1941, 447–449

4 Douglass, H R "Means of Predicting Scholastic Success in the College of Dentistry at the University of Minnesota." *Proceedings of the American Association of Dental Schools*, XIV (1937), 422–431

5. Douglass, H. R. "Factors Associated with Scholastic Success in the School of Dentistry at the University of Minnesota." *Proceedings of the American Association of Dental Schools*, XV (1938), 172–179.

6. Freeman, H. H. and Smith, R. V. "A Report on Aptitude Testing at the University of Iowa." *Proceedings of the American Association of Dental Schools*, XII (1935), 214–228

7. Graver, Harold A. "Report of the Committee on Aptitude Testing." *Proceedings of the American Association of Dental Schools*, XVII (1940), 105–113

8. Graver, Harold A. "Factors in Dental Aptitude." *Proceedings of the American Association of Dental Schools*, XIX (1942), 252–257.

9. Kuder, G Frederic. *Manual for the Preference Record* Chicago Test Service Division, Science Research Associates (pp. 11–13 especially)

10. Kuder, G. Frederic. "The Stability of Preference Items." *The Journal of Social Psychology*, XIX (1939), 41–50.

11 MacQuarrie, T. W. *MacQuarrie Test for Mechanical Ability* (Manual of Directions) Los Angeles. California Test Bureau (pp. 1–2 especially)

12. Moss, F. A. "Desirability of Choosing and Means of Selecting Students." *Proceedings of the Ninth Annual Meeting of the American Association of Dental Schools*, 1932

13. Moss, F. A. "Annual Reports of the Secretary of the Committee on Aptitude Tests for Medical Students." *Journal of the Association of American Medical Colleges*, Sept., 1931, May, 1932; Mar., 1934; Jan., 1935

14 Stoddard, G. D "Some Factors Related to Success in the Study of Dentistry." *Proceedings of the American Association of Dental Schools*, VIII (1931), 56–65.

15. Thompson, Claude E. "Motor and Mechanical Abilities in Professional Schools." *Journal of Applied Psychology*, XXVI (1942), 24–39

16. Tiegs, E. W., Clark, W W and Thorpe, L. P. *Manual of Directions, California Test of Personality—Adult Series, 1940.* Los Angeles: California Test Bureau.

17. Traxler, Arthur E. and McCall, William C. "Some Data on the Kuder Preference Record." *Educational and Psychological Measurement*, III (1941), 253–268.

THE WORD-DEXTERITY TEST, A BETTER MEASURE OF COLLEGE APTITUDE

SHAILER PETERSON
The University of Chicago

THE purpose of this article is to describe a measuring instrument which, as an aptitude test, compares favorably with examinations commonly given during Freshman Week. This test has been tried at junior-high-school, senior-high-school and college level and has proved itself a valuable predictor of school and course grades

The origin of this examination dates from work in remedial reading carried on at the University of Oregon in 1933, in cooperation with the late Dr. DeBusk. Experimentation indicated that there was a distinct improvement in working vocabulary as soon as the student could see the carry-over in word meaning from one word to another. While this seemed to be particularly marked in the field of the natural sciences, other work has indicated that this ability aids students in other areas. The first instruments were not tests but instead were primarily teaching devices and intended for students who required remedial assistance to improve their school marks. New instructional devices were prepared at Lebanon High School and at the University of Oregon High School At this time, the author described these first efforts in *The English Journal.*[1] Later at the University of Minnesota, with the encouragement of Dr. Alvin C. Eurich, Dr. Guy Bond, and Dr. Palmer O Johnson, more experimental work was conducted on this same project. Still more recently at South Dakota State College there was further opportunity for the author to observe the value of the *Word-Dexterity Test* that had been developed.

---

[1] Peterson, Shailer A "Teaching the Special Vocabularies." *The English Journal (College Edition)*, XXV (1934), 53–56.

The main objectives of the test in its present form are to determine to what extent the student knows the meaning of certain suffixes and prefixes in common use and also to determine if he can transplant the meaning of a suffix or prefix found in one word, whose meaning is known, to another word in which the same suffix or prefix is found From an examination of the test items illustrated in this article, the reader can understand that while pure memory of words, suffixes, and prefixes will assist the student in securing a high score, still when the range of word difficulty is at the proper level, the test becomes essentially a problem for the student to demonstrate his "dexterity" at manipulating word parts and word meanings. Mechanically, this also becomes an "analogy" type test or one which tests for configuration and pattern.

From an examination of the following directions and sample test items, the reader will be able to understand the construction of the entire test. It is a power test rather than a speed test and the fifty-item test described here could be administered to senior-high-school or college students in 40 minutes. The words used in the examination were chosen after consideration of their Thorndike word count These words, in addition to having a fairly extended range of difficulty, all employed suffixes and prefixes which in turn were to be found in at least four or five very common words as revealed by their Thorndike word count.

Directions

In this test, there are many scientific words that you already know Of the others, you will find that in many cases you will be able to discover their meaning as you proceed.

In the example, the *word part* SUB is considered. There are many words in our English language containing SUB One of these words, SUBMARINE, has been printed on the line alongside. There is a space for two other words containing this same *word part*, SUB, and in this example, the two words, SUBWAY and SUBNORMAL have been written in.

Example: SUB submarine subway subnormal

Each of the three words, SUBMARINE, SUBWAY, and SUBNORMAL, contain the same common *part* SUB The first word

refers to an *under*-water boat, the second word refers to an *under*ground railway; while the last word refers to things or conditions that are *below* or *under* normal All three of these words not only have the same *word part*, SUB, but they all have the same meaning, "*under*"

On another line beside each of the sections, there is a group of *definitions*, one of which is correct. You are to select the correct definition and place its letter in the space in the margin. As you will see in the following part of the example, definition "B" best describes the *word part*, SUB

Example: B SUB (A) good quality, excellent; (B) beneath, below, underneath; (C) mathematical, scientific; (D) slow, delayed; (E) similar in appearance, alike, a duplicate

While there is a space in each section for you to write only two additional words, you may be able to think of more than two. In some cases you may have difficulty in thinking of more than one additional one The more words that you can think of, the easier it will be to decide what the true meaning is, for as in the case of the words submarine, subway, and subnormal, each word is alike in two respects These words have the similar *word part* SUB and also have the similar meaning, "under"

You will find that this is true of other words also. Whether you can think of five new words or only one or two, try to decide on the best definition for the *word part* under consideration.

DON'T WASTE TIME. DO FIRST THOSE ITEMS WHICH ARE EASIEST FOR YOU.

Test Items

______ 3. MIS misunderstand ______ ______

(A) wrong; (B) small, tiny, petty, (C) excuse, reason; (D) denotes feminine gender; (E) poor, inferior

______ 4 CO cooperate ______ ______

(A) with, together; (B) two, (C) without stopping, (D) aid, help; (E) easy, ready

______ 5. IST chemist ______ ______

(A) science, logical study, investigation, (B) schooling, training, preparation; (C) forming agent nouns; (D) helping, assisting, aiding; (E) including, possessing, having.

______ 6. LOGY mineralogy ______ ______

(A) weather, climate, temperature, (B) mining; (C) prediction, estimation, (D) discourse, theory, doctrine; (E) tired, weary, slow

______ 10 GRAPH biography ______ ______

(A) living things, life, (B) important, distinguished; (C) soil, ground, rocks; (D) drawing, writing; (E) a study, a science

______ 11 ANTI antislavery ______ ______

(A) treatment, medical aid, (B) old, historical, not used; (C) in front of, before, preceding; (D) opposite, against, instead of; (E) a study, a science.

______ 32. SCRIB transcribe ______ ______

(A) radio, phonograph; (B) circular, round, (C) write; (D) duplicate, (E) voice, vocal

______ 34. POLY polytechnic ______ ______

(A) arithmetic, numbers; (B) many, much, often; (C) few, not many, (D) figures, diagrams; (E) college, school, study.

______ 39 GEN genealogy ______ ______

(A) science, study; (B) birth, born, descent; (C) related, corresponding; (D) electrical; (E) Bible, Biblical

______ 44. PATH pathology ______ ______

(A) scientific, (B) anger; (C) suffering; disease, (D) judgment, estimation, (E) germs, bacteria.

______ 47. DECI decimal ______ ______

(A) arithmetic process, multiplication, etc.; (B) half; (C) ten, ten times, (D) almost, nearly, quite, (E) accurate, scientific, sound.

______ 48. DUCT ductile ______ ______

(A) save, preserve; (B) stretch, lengthen, (C) press, flatten, squeeze; (D) able to be attracted, attractive, (E) lead, direct, guide.

______ 49. ANTE antedate ______ ______

(A) outside, beyond; (B) behind, beside, (C) in front of, preceding, before; (D) old, historical, not used; (E) not, never, none.

In these items, the student is asked to write down a group of other words each of which contains some of the same prefixes and suffixes. This is to assist him in assigning word meaning to the word parts in the unknown word. The test is scored,

however, *on the basis of correct responses to the word's meaning* and ordinarily no attention is paid to the particular "assist" words that he writes down. If this examination were administered in a situation where attention was to be given to remedial work, then the character of these "assist" words would be important.

Table 1 shows the product-moment correlation coefficients between different variates in a college group of nearly three hundred students From this table, it is evident that the *Word Dexterity Test* predicts grade-point average, English grades, mathematics grades, or chemistry grades better than most of the other examinations with which it was compared. Grades themselves in some subjects were better predictors of total grade-point average, but this can be explained by the fact that the individual course grades were themselves a part of the grand average.

TABLE 1

Correlation Coefficients between Different Variates in College Group

	Peterson Word-Dexterity	Total A.C.E.	Q Score A.C.E.	L Score A.C.E.	Speed Reading Comprehension	Vocabulary Reading Comprehension	Level of Reading Comprehension	Total Grade Average	English Grades	Mathematics Grades	Chemistry Grades
Peterson Word-Dexterity		.44	.43	.63	.52	.62	.50	.56	.58	.33	.49
Total A. C. E.	.44		.82	.89	[illegible]	[illegible]	[illegible]	.42	[illegible]	.26	.42
Q Score A. C. E.	.42	.82		.52		[illegible]	[illegible]	.31	[illegible]	.25	.43
L Score A. C. E.	.63	.89	.52		[illegible]	[illegible]	[illegible]	.42	[illegible]	.16	.39
Speed Reading Comprehension	.52	[illegible]	[illegible]	[illegible]	[illegible]	.68	.83	.42	.50	[illegible]	[illegible]
Vocabulary Reading Comprehension	.62		[illegible]	[illegible]	.68	[illegible]	.65	.40	.43	[illegible]	[illegible]
Level of Reading Comprehension	.50				.83	.65		.39	.44	[illegible]	[illegible]
Total Grade Average	.56	.43	.31	.42	.42	.40	.39	[illegible]	.79	.82	.68
English Grades	.58			[illegible]	.50	.43	.44	.79	[illegible]	[illegible]	.43
Mathematics Grades	.33	.26	.25	.16		[illegible]	[illegible]	.82	[illegible]	[illegible]	.43
Chemistry Grades	.49	.42	.62	.39	[illegible]	[illegible]	[illegible]	.68	[illegible]	.43	[illegible]

At the high-school level, the product-moment correlation between the Word Dexterity Examination and I.Q. was +.58

and with mental age, it was + .70. The point biserial correlation[2] with years of science was + .20 and with years of foreign language was + .39. The biserial correlation with science grades was + .52.

While the scores for the 7th- and 8th-grade students are considerably lower than those for any of the other groups, it is interesting to observe that there is considerable overlapping of the scores for the upper grades. Figure 1 illustrates the percentile rank curves for the seven grade levels.

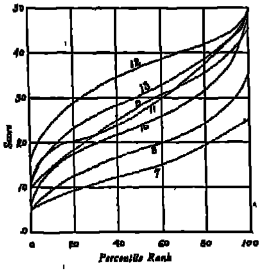


FIGURE 1
Percentile Rank Curves for Peterson Word Dexterity Test

The estimated reliability of the forty-minute Word-Dexterity Test was + .90 to + .93 depending upon the groups. These computations were made by the Hoyt method[3] and by Formula 20 of the Kuder-Richardson method,[4] each of which gives underestimates of the true reliability. In order to evaluate the function of each of the items in the examination, the item-test

correlation for each of the items was computed both on the examination administered at the high-school level and on the one administered at the college level. Table 2 provides a frequency distribution of the fifty items according to the correlations that they were found to have with the total test, a measure of item validity when the total test is the criterion measure. The high item-test correlations for all items in the examination must be interpreted to mean that there is a commonality of purpose and function for all items in the examination and that whatever is being measured by the entire examination is being contributed to materially by each item in the examination.

TABLE 2
Distribution of Item-Test Correlations Based upon Top and Bottom Quarters

Item-Test Correlation	Frequency of Items	
	High-School Group	College Group
+ .8 to .9	3	1
.7 to .8	10	3
.6 to .7	16	12
.5 to .6	8	13
.4 to .5	10	6
.3 to .4	2	9
.2 to .3	0	5
.1 to .2	1	1
.0 to .1	0	0
Median Value of r	+ .63	+ .53

Sample copies of the author's *Word-Dexterity Test* are available directly from him and permission to duplicate this examination can be secured.

---

[2] The point biserial r does not assume a normal distribution of the variate and is in these instances a better estimate than the biserial r that does assume a normal distribution.

[3] Hoyt, Cyril. "The Reliability Estimated by Analysis of Variance." *Psychometrika*, 4 (1941), 153–160.

[4] Richardson, M. W. and Kuder, G. F. "The Calculation of Test Reliability Coefficients Based on the Method of Rational Equivalence." *Journal of Educational Psychology*, XXX (1939), 682–683.

A STUDY OF THE KUDER PREFERENCE RECORD

DANIEL J. BOLANOVICH AND CHARLES H. GOODMAN
Radio Corporation of America

AT the beginning of the war, one of the major problems that faced industry was the need for highly trained technical personnel. The Radio Corporation of America, like many other companies faced with a shortage of engineers, planned to meet this need with a short intensive training program for young women. The program sponsored by RCA was a ten-month course in the theory and practice of electronic engineering[1] and was given at Purdue University.

Because of the importance of obtaining a maximum number of technical personnel who would be successful in the course, psychological tests were used as part of the selection process. The findings which were obtained as a result of the use of these psychological tests will be published at a later date.

Three hundred applicants from various colleges of thirty-six states and from RCA's six manufacturing plants applied for the course. Of this number eighty-six were finally selected. The girls selected were called RCA Cadettes. Their ages ranged from eighteen to twenty-nine years, with a median age of nineteen years. Five Cadettes had had no previous college experience, sixty-five of the group had either one or two years of college work when they entered the course, and one Cadette had had graduate work.

The positions which the Cadettes were expected to fill were varied and included such activities as writing instruction manuals, design, drafting, statistical analysis, laboratory assembly, and equipment testing. Because of the variety of jobs open to the Cadettes it was felt that it would be helpful in placing them

if their interests could be determined. One month after the Cadettes had been at Purdue University the Kuder *Preference Record* was administered to them. The reasons for using the Kuder were: (1) it offered interest areas which appeared related to the jobs that were to be filled; (2) it offered possibilities of determining the preferences of the Cadettes for placement, (3) it offered an opportunity to examine the Kuder as a possible selection device for future training programs, and (4) it was easy to score.

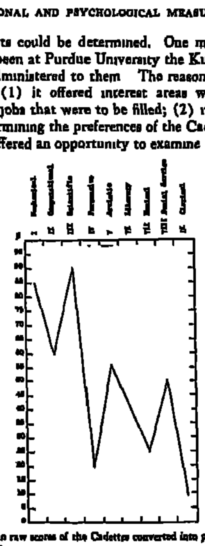


FIG. 1. Median raw scores of the Cadettes converted into percentile equivalents on the Kuder norms.

The results reported in this paper are based upon the scores obtained from the sixty-six Cadettes for whom there were complete data. Upon completion of the scoring of the *Preference Record*, an individual profile was constructed for each Cadette. In order to obtain a graphic profile that would be representative of the Cadettes as a group the median scores were calculated for the Cadettes on each category of the *Preference Record.*

---

[1] The courses given in the training program were: Mathematics, Shop, Drawing, A-c, D-c, Radio Manufacturing, Communications, Radio, Measurements, Electronics.

Figure 1 shows the profile constructed by plotting these median scores after converting them into the percentile equivalents given by Kuder in his norms for college freshmen [3] It is interesting to note from Figure 1 that the Cadettes as a group stand at the 84th percentile in the Mechanical preference area and at the 92nd percentile in the Scientific preference area There is, of course, the possibility that the Cadettes' responses may have been influenced by their desire to appear highly interested in scientific and mechanical pursuits because of the nature of the course they were taking. It should be remembered, however, that the Kuder *Preference Record* played no part in the selection process of the Cadettes. The Cadettes appear to be similar to the norm population in Computational, Artistic, and Social Service preferences. They are, however, considerably below the median of the norm population in Persuasive, Clerical, and Musical preferences.

The graphs in Figure 2 show the percentage distributions of the Cadettes' scores on each of the preference categories The norm graph shown was constructed in order to facilitate a visual comparison of the percentage distributions found for the Cadettes on each of the preference categories. The base lines of each graph have been divided at interval points corresponding to the 10th, 25th, 50th, 75th, and 90th percentiles for the norm group Each graph can then be compared to the norm graph.

Figure 2 appears to support the evidence of Figure 1 of the selection of a group particularly interested in the Scientific and Mechanical categories, a finding similar to that obtained by Goodman[4] for engineering students. There are, however, a few Cadettes below the norm medians on these keys The distribution of the percentages on the Clerical graph is the reverse of those on the Mechanical and Scientific categories Forty-four per cent of the Cadettes fall below the 10th percentile and only four per cent are above the 75th percentile. The majority of the Cadettes fall below the 25th percentile in Persuasive preference, although there are five per cent above the 90th percentile who might fit into jobs where considerable contact with others is necessary, such as quality control inspectors. The Cadettes tend to fall in the middle ranges in Computational Preferences

---

[3] Norms for women only were not available at the time this study was made, Use of norms for high-school girls, which have since been published, would not change the conclusions reached in this paper

[4] Cf Goodman, C. H "A Comparison of the Interests and Personality Traits of Engineers and Liberal Arts Students" *The Journal of Applied Psychology,* XXVI (1942), 721–737

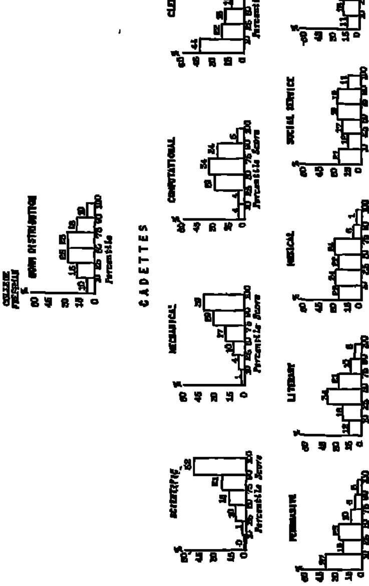


FIG. 2. Kuder *Preference Record* Distributions of RCA Engineering Cadettes.

These graphs were useful in helping to determine final placement of the Cadettes. The Cadettes in the extreme ten per cent groups of each Preference category were given particular attention for possible placement in jobs related to their preferences. For example, some of the jobs to be filled involved mechanical work, scientific work, clerical work and literary work The histograms in Figure 2 were also used in counseling the Cadettes and aiding them in the interpretation of their own individual profiles

To evaluate the possibilities of the *Preference Record* as a selection tool for similar programs, correlations of the Cadettes' Preference scores with their final grade averages for the entire course were obtained. It was found that first- and second-semester grade averages of the Cadettes correlated .81, which would indicate fairly close agreement between the grades of the first and second semesters. As a result total grade averages were taken as the criterion. Table 1 gives the frequency distribution of grade averages, and shows a range from 3 0 to 5.9. The grades for each subject were given in whole numbers ranging from 2 to 6. Grade averages were obtained by multiplying each subject grade by the number of credit hours and dividing the total for all subjects by the total number of hours for the course

TABLE 1

Distribution of Cadettes' Total Grade Averages

Total Grade Averages	Number of Cadettes
5.7–5.9	3
5.4–5.6	5
5 1–5.3	5
4.8–5.0	12
4.5–4 7	9
4.2–4 4	11
3.9–4 1	10
3.6–3.8	5
3 3–3 5	6
3.0–3.2	2
	N=66

Table 2 shows the correlations of grade averages with the various Preference scores. Since none of the correlations attain statistical significance[4] (even at the five per cent level) the Preference scores cannot be considered as predictive of success in the Cadette training course.

For further clarification of the relationship between Preference scores and success in the training course, an analysis was made of the scores of three sub-groups—(1) the most successful students, (2) the least successful students, and (3) those who terminated their training before the course was completed Figure 3 graphically shows the median scores converted into their percentile equivalents on the Kuder norms of 16 Cadettes with grade averages of 4.9 or above, 13 Cadettes with grade averages of 3.8 and below, and the 8 Cadettes who dropped out of the course for reasons other than those attributed to illness. For groups as small as these, the proximity of the curves for the most successful and least successful Cadettes who completed the course appears pronounced. Figure 3 seems to bear out the low correlations of Table 2 The largest differences between

TABLE 2

Correlations of the Kuder Preference Record Scores with Total Grade Averages

Kuder Key		Correlation with Total Grade Averages
I	Mechanical	.09
II	Computational	.18
III.	Scientific	10
IV	Persuasive	15
V	Artistic	.08
VI.	Literary	.02
VII	Musical	–.02
VIII.	Social Service	–.03
IX.	Clerical	– 14

---

[4] Lindquist, E F *Statistical Analysis in Educational Research.* New York: Houghton Mifflin Co., 1940. P 212.

the most successful and the least successful Cadettes appear to be on the Computational, Artistic, Musical, and Social Service categories

The profile of the terminated Cadettes appears to deviate from the patterns of the groups who completed the course. The terminated Cadettes have lower median percentile scores in Mechanical, Computational, and Scientific preferences, and higher median percentile scores in the Persuasive, Literary, and Social Service categories.

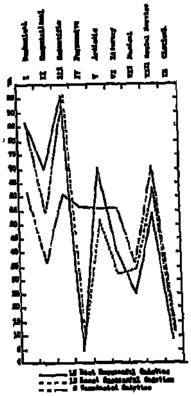


FIG. 3 Comparison of Median Raw Scores for the Most Successful, Least Successful and Terminated Cadettes Converted into Percentile Equivalents on the Kuder Norms.

In order to determine the significance of the differences shown in Figure 3, *t* ratios were computed using Student's technique for small sample distributions Table 3 shows the median

raw scores for the three sub-groups on each of the Kuder preference scales The *t* ratios obtained for each group, when compared with the other two groups and the probability values for these *t* ratios, are shown in Table 4

Accepting Fisher's criterion of a 5% probability as indicative of a significant difference, it will be found that there are only five such *P* values in Table 4 The five instances of significant differences are

TABLE 3

Median Raw Scores of the Most Successful Least Successful, and Terminated Cadettes on the Kuder Preference Record

Groups	I Mechanical	II Computational	III Scientific	IV Persuasive	V Artistic	VI Literary	VII Musical	VIII Social Service	IX Clerical
Most Successful (N=16)	75.5	41.0	78.0	48.5	58.0	51.5	18.5	73.0	37.5
Least Successful (N=13)	74.0	31.0	76.0	55.0	50.0	50.0	23.0	77.0	[illegible]
Terminated (N= 8)	59.5	24.5	55.5	74.5	51.5	60.5	22.0	81.0	40.5

Computational—	Most successful Cadettes score significantly higher than the least successful Cadettes.
Computational—	Most successful Cadettes score significantly higher than the terminated Cadettes.
Scientific—	Most successful Cadettes score significantly higher than terminated Cadettes.
Persuasive—	Terminated Cadettes score significantly higher than most successful Cadettes.
Persuasive—	Terminated Cadettes score significantly higher than least successful Cadettes.

The results indicate that the most successful Cadettes are differentiated from the least successful Cadettes on the Computational key only The most successful Cadettes are differentiated from the terminated Cadettes on the Computational and

TABLE 4

Comparison of the t Values and the Probability of Significant Differences Among the Most Successful, Least Successful and Terminated Cadettes on the Kuder Preference Scales

Groups Compared	I Mechanical		II Computational		III Scientific		IV Persuasive		V Artistic		VI Literary		VII Musical		VIII Social Service		IX Clerical	
	t	*P*	*t*	*P*	*t*	*P*	*t*	*P*	*t*	*P*	*t*	*P*	*t*	*P*	*t*	*P*	*t*	*P*
Most successful Terminated	1.52	.20	2.20	.05	1.98	.05	−1.23	.01	.48	.60	−.98	.30	−.27	.80	−.86	.40	−.53	.60
Least successful Terminated	1.32	.20	.98	.30	1.87	.10	−2.36	.02	−.15	.90	−1.24	.20	.00	1.00	−.16	.70	−.07	.90
Most successful Least successful	.21	.80	2.40	.02	.23	.80	−1.04	.30	.82	.40	.48	.60	−.84	.40	−.40	.70	−.19	.70

Scientific keys The terminated group is differentiated from the most successful and least successful Cadettes by higher scores on the Persuasive scale. On the whole the group of terminated Cadettes appears to approximate more nearly the pattern of preferences of Kuder's college freshmen

The above differences provide the only evidence that may indicate any possible utility of the Kuder *Preference Record* as a selection device for future applicants in a training course such as the one described.

Summary

This study has presented data based upon the Kuder *Preference Record* scores of 66 Cadettes enrolled in a training program for electrical engineering aides. Investigation was made of the possibilities of the Kuder *Preference Record* as a selection device and its usefulness as a counseling and placement tool The findings show that the Cadettes, as a group, differed from the college freshmen norm group, having particularly strong preferences in Mechanical and Scientific pursuits There was a pronounced lack of preferences in Persuasive, Clerical, and Musical endeavors. Score distributions showed, however, wide differences among the Cadettes on each scale. There were, however, a few Cadettes who showed a lack of preferences in Scientific and Mechanical fields while a few ranked very high in Clerical, Computational, Literary, and Persuasive preferences. These distributions, together with the individual profiles, provided some indications for counseling and placing the Cadettes.

The scales of the *Preference Record* showed low correlations with final grade averages for the course. The highest correlation was .18 for the Computational key with final grade average.

The profiles of the most successful Cadettes and the least successful were similar. The group profile of the terminated Cadettes approached that of the average college freshmen. These terminated Cadettes made significantly lower scores on the Scientific and Computational preferences than the most successful Cadettes. The terminated Cadettes made significantly higher scores than the most successful and least successful Cadettes on the Persuasive scale. The most successful Cadettes made significantly higher scores than the least successful Cadettes on the Computational key

The following conclusions appear to be justified·

(1) On the basis of the correlations with total grade averages, the Kuder *Preference Record* does not appear to be a promising selection device for predicting course achievement of female engineering Cadettes It does, however, show some promise as a device for eliminating those who would be likely to drop out before completion of the course

(2) *The Kuder scores afford some indications that can be* helpful in counseling and placement, especially in a situation where there is a variety of job openings

---

DEVELOPING A SERVICE RATING SYSTEM

IRENEUS S SMITH
Civil Service Commission of San Francisco

THE "San Francisco System" of service rating was put into effect by the Civil Service Commission of San Francisco on July 1, 1944, for probationary employees, and shortly will be applied to nearly 15,000 employees of the County and City of San Francisco. Before the system was developed and adopted, a comprehensive survey was made of various systems in use. This paper is concerned with a presentation of the findings of our survey and a description of the approach finally taken in developing the "San Francisco System."

Our research took many months, covering most of the major federal, state, and municipal personnel jurisdictions throughout the United States and Canada We received the finest sort of co-operation almost uniformly, and in our analysis of the different systems in use, criticism is not intended We analyzed each in relation to the specific problems facing us in developing a rating system that would be of the most value for our own local jurisdiction

The problem of rating employee performance is still in the experimental stage and, up to comparatively recent times, ratings were considered more or less unreliable Personnel agencies have lately devoted a very large amount of research to the problem, while leading personnel writers and publications have exhaustively gone into the field

From San Francisco's viewpoint, the purpose of inaugurating a service rating system was to develop a means of attracting more capable workers into public service as a career, raising the caliber of city personnel, in line with the merit system trend to further eliminate the "spoils system" from public service. Such a service rating system was authorized by our city charter.

Our national survey showed that service rating has proved of real value in raising civil service standards It is desired generally by competent employees since recognition may be given to faithfulness and merit as factors in promotional examinations as well as in eligibility for salary increases, transfers, leaves of absence, and other civil service privileges.

From the appointing officer's viewpoint, service rating provides an impartial yardstick by which he can measure employee performance, and also provides a definite incentive by which the majority of employees seek improvement, to that extent bettering the performance of his department and at the same time simplifying his personnel problem. From this definite national consensus, we felt assured in following through with development and inauguration of our own system

We also found that regardless of the form of the rating device there are, in the opinion of administrators generally, certain requisites to the success of any employee rating plan·

First· The supervisors or executives who are asked to rate their subordinates must be sold on the idea. This result may be brought about by a series of conferences with such supervisors conducted by officials of the personnel department.

Second The employees themselves *should be sold on the* idea. When merit ratings are properly sold, investigators report that a large percentage of employees like to be rated and will give complete co-operation to any rating plan that they *consider to* be fair.

Third· The supervisors must be trained to rate those under them. One of the principal weaknesses of rating systems appears to be the lack of uniformity of ratings between different departments or bureaus In one department, for example, a supervisor may set very low standards of performance and rate 85% of his employees "excellent" and the remaining 15% "very good" In another department the supervisor may set a very much higher standard of performance and rate 85% of his employees good and distribute the other 15% anywhere along the scale from excellent to unsatisfactory. In attempting to solve the problem of maintaining reasonable uniformity of service rating standards some jurisdictions—notably, the United States

Housing Authority, the Tennessee Valley Authority, and the New York State service—have adopted a plan based on the idea that the categories of "excellent" and "unsatisfactory" should be reserved for those employees who distinguish themselves from their fellow workers unmistakably, to either the advantage or the disadvantage of the service, and when such ratings are given, the supervisor is required to cite on the back of the rating form, or on a special form, concrete evidence of such service

Fourth: The rater should take the ratee into his confidence, show him his ratings, and discuss them with him When the employee knows what traits or qualities are judged to be unsatisfactory, his normal reaction would be to seek self-improvement Especially will this be the case when ratings are used as a factor in promotional examinations, and also in determining the order of layoff and reemployment, and eligibility for salary increases, transfers, leaves of absence, and other privileges. How the employee does his work must be determined in most cases by the employee's immediate supervisors One of the problems therefore is to secure reliable ratings from various supervisory officers. In this respect there are two schools of thought. *One maintains that the actual assignment of a* rating should be made by the supervisors who are directly in charge of the employees The other and more recent school confines the supervisor to the reporting of significant behaviors or activities, *relying for the actual valuation upon some* mechanical scoring system administered by the central personnel staff

In devising a rating system the two principal technical problems have to do with the selection of factors to be rated and the comparative evaluation of such factors. Authorities generally are agreed that the objectivity and reliability of ratings increase as the factors involved become more specific. The chief fault to be found with rating plans in the past is that such plans provided for over-all ratings by the supervisor. Present practice is opposed to reliance upon such highly subjective estimates. It prefers, overwhelmingly, ratings which are objective. The maximum degree of objectivity is approached if the attention of the supervisor is focused on specific job behaviors or activities. In addition, each item to be reported or evaluated should be clearly described in simple, unambiguous terms and as concretely as possible

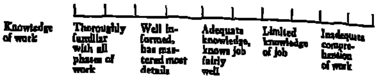
The point should be stressed, however, that no matter how carefully factors have been selected and no matter how carefully supervisors observe and report the behaviors of their employees, the whole system will break down unless the actual rating based on these is sound Almost without exception, in the opinion of Mosher and Kingsley, the devices that have been developed to evaluate employee services have failed to provide an adequate measuring instrument.

One method of rating the efficiency of employees, especially in industry, is through the use of production records. Such positions as stenographer, typist, file clerk, machine operator, or copyist lend themselves to unit measurement. In some of the larger federal establishments, such as the Farm Credit Association and the Federal Reserve Bank of New York, the number of pages typed, the number of errors, and the general appearance of the work are taken into consideration by supervisors when typists are rated.

Another method is through the use of periodic tests. This method has been extensively employed in some of the federal bureaus for rating the efficiency of their employees. In the Department of Justice, for example, the service ratings of stenographers and typists depend in part upon their standing in periodic speed tests, while performance on a periodic sorting test is one basis for determining the rating of postal clerks This method of rating, however, is adapted only to routine and repetitive jobs.

The method in most general use is that of rating schedules. A description of some of the schedules in use in other jurisdictions follows

1. *The Graphic Rating Scale.* This, according to many personnel experts, is the most popular rating method in use, being widely employed in private concerns as well as by a number of public personnel administrations It consists essentially of two elements: (1) a list of traits or activities arrived at by an analysis of factors leading to success or making for failure on the job, (2) various descriptive phrases or adjectives denoting the several degrees of a particular activity or trait. The form of the device is as follows.


Knowledge of work
Thoroughly familiar with all phases of work
Well informed, has mastered most details
Adequate knowledge, knows job fairly well
Limited knowledge of job
Inadequate comprehension of work

The descriptive phrases serve as a guide to the rater who is instructed to place a check mark along the line in one of the 10 boxes above the phrase or between the phrases at the point which, in the rater's opinion, represents the degree of the particular quality possessed by the ratee.

This method is open to the objection that there is no standard unit of measurement involved and the device is not really a scale at all. It is a schedule the various items of which are arbitrarily weighted and given a numerical value through the use of a scoring stencil The device is scored as a rule by the application of a ten-point stencil to the straight line.

2. *The Probst Service Rating System.* This system, developed by J B. Probst, the Chief Examiner of the St Paul Civil Service Commission, in 1928 was in the opinion of personnel administrators the only system which up to that time showed any real marks of merit. It has been tried out in a number of civil service jurisdictions including the Cincinnati, Detroit, Los Angeles, and California State public services. Mr. Probst discarded the scale idea previously mentioned and substituted for it a list of about 100 characteristics or modes of behavior The rater checks only those items which are known to describe the ratee; the rest he leaves blank. In addition, he is not required to measure relative degrees of a quality in the ratee According to its author the scheme is so designed that failure on the part of the scoring official to correctly and conscientiously check the employee's traits can be ascertained at a glance. It has the added advantage that elaborate instructions to reporting officers are unnecessary.

The actual evaluation of employee service is made by the use of a mechanical scoring system administered by the central personnel office. This scoring system has been one of the principal bases of criticism of the plan because of the difficulty in understanding it.

3. *California Report of Performance Plan.* In 1938 the California State Personnel Board abandoned the Probst System then in use for one which, in the opinion of the Board, more satisfactorily and adequately measured the performance of State employees. A separate report sheet for different types of work containing a particular combination of work characteristics was developed At the present time 45 different report forms each containing a separate list of factors to be rated are employed. There are five possible gradations of markings for each item, represented by five columns These columns are defined in terms of the extent to which an item is characteristic of the work of an employee. The individual in highest authority who is in intimate contact with the work of the employee acts as the "reporting" officer. After the reporting officer has prepared the report he must review it and discuss it with the employee, who is also given a copy of the report. The original is sent to the personnel agency for scoring. An appeal board has likewise been set up to hear and decide cases in which the supervisor and the employee disagree on the markings of a particular report.

4. *The Los Angeles City Schools* recently adopted a performance report system of the trait-rating type where four separate forms are used. A unique feature of this plan is that it was developed by representatives of employee groups and organizations. These groups studied and discussed the problem for about a year before submitting the final plan to the personnel board. The forms are so compiled that they may be machine scored. These two features, and especially the first, might well be given favorable consideration if and when a rating system for all employees is put into effect Employee support would thus be assured, without which, it is claimed, no plan can succeed.

5. *The San Francisco public schools* are at present using a rating form for non-certified personnel. It consists of a three-step rating schedule in which three degrees of each trait corresponding to excellent, good, and poor are identified by appropriate descriptive phrases arranged in three columns Such a performance report is open to the objection that the ratings are too broad or, in other words, that they do not give a sufficiently fine graduation of the relative excellence of the employees.

6. *The Home Owners Loan Corporation* utilizes a service rating form that consists of three traits which are rated directly by the rating officer on the basis of excellent, very good, good, fair, or unsatisfactory. This method is open to objections, among them these: the traits to be rated are too few to yield a comprehensive picture of the employee's ability, also, the traits are compound ones and require an over-all rating of more than one type of activity.

7 *The City of St. Louis, Department of Personnel,* very recently adopted a service rating plan according to which employees are rated on eight specific traits and an over-all evaluation of work performance. These nine measures are felt to include all factors necessary to arrive at a comprehensive evaluation of employee performance and to apply to all types of positions. This plan, like that of the H O L C., is open to the criticism that the ratings are too subjective since the traits are rated by two different supervisors who place a check mark in one of five columns headed excellent, very good, good, fair, and poor The subjectivity of the ratings is lessened to some extent, however, by a provision whereby the supervisor is required, when a rating of either excellent or unsatisfactory on the over-all evaluation is given, to furnish substantiating evidence of the employee's superior or inferior performance on the job.

8 *The Detroit municipal service rating report* consists of a three-step rating schedule in which degrees of poor, satisfactory, and above average of 25 traits are rated objectively with the aid of descriptive phrases which identify the various degrees of the trait Ten additional traits are rated where the employee is in supervisory charge of other employees' work. It appears to have considerable merit. The number of traits rated objec-

tively is sufficiently large to give a comprehensive and reliable picture of the employee's worth. The one criticism that might be leveled at the method of rating is that only three degrees of the trait are rated in place of the conventional five-step rating schedule.

9. *The Minnesota State* report of employee performance is an adaptation of the California State employee performance form. It consists of a large number of traits (about 60 in all) which may be measured objectively in terms of the frequency with which these traits can be observed. The traits are checked in one of five columns. Different forms are used to rate different groups of employees. Complete instructions for use of the rating form are contained on the reverse side of the form. The large number of traits to be rated makes the scale rather cumbersome; it has the merit, however, of giving a very comprehensive picture of employee worth.

10. *Saginaw, Michigan,* uses an employee service report of the graphic rating scale type. Eight traits are rated objectively with the aid of phrases descriptive of degrees of the trait [illegible] worth in the opinion of the rater. The form has the merit of extreme simplicity and allows the rater wide latitude in his appraisal of the degree of each trait possessed by the ratee.

11. A distinct departure from the conventional rating schedule in general use is the "*Employee Guidance Sheet,*" as it is called, recently developed by the *Alabama State Personnel Department.* The horizontal scale has been discarded in favor of a check-list arrangement. The five descriptive phrases identifying the various degrees of the ten traits to be rated are couched in language intended to help and encourage the employee rather than to report findings in a coldly impersonal and sometimes quite blunt manner. The following sample will illustrate the effort that was made to humanize the report and stimulate the employee to greater effort.

Usual Form

Quantity of work.
( ) Unusually high output.
( ) High output.
( ) Normal output.
( ) Limited output.
( ) Insufficient output, unsatisfactory

Alabama State Form

Quantity of work
Just a friendly suggestion.
( ) Exceptionally high output. Keep it up.
( ) Better than average. Good going.
( ) Meeting our requirements.
( ) You could do more. Try harder.
( ) You could do a lot more. Try much harder

Other features include a statement in which the reporting officer indicates whether the duties of the position have changed since the last rating period and also a scale on which the department head indicates his opinion of the supervisor's ability as a rater. All in all, this rating form seemed by far the best of any studied by us. It combined all the features of an exceptionally well developed rating system, including simplicity, ease in scoring and rating, and high objectivity and reliability. In addition, the language employed in the rating scale was so chosen that the employee would feel that the criticism was offered in a helpful mood and in consequence should have a strong desire to improve his performance if it were below normal.

The foregoing discussion is devoted to a review on a very limited scale of representative rating plans in use. None of these plans are restricted to the rating of the probationary period alone. In inaugurating the San Francisco employee rating plan, therefore, it was felt desirable, in order to acquaint employees and employers with the idea gradually, to limit it to probationers first. The San Francisco charter provides that "at any time during the probationary period the appointing officer may terminate the appointment." Since power of dismissal is vested in the appointing officer and since his decision is final, we felt that simplicity should be the determining characteristic of any rating plan for such employees. Keeping in mind the objectives that should be inherent in every good rating device, we attempted to develop a plan which included the best features of those reviewed, with particular emphasis on the following points:

1. The development of a uniform report sheet containing traits which will fit all types of work.

2. Clear and concise wording of the descriptive phrases which measure the degree of each trait.

3. Phrasing of steps so as to compel supervisors to give careful consideration to their markings.

4. A sound and easily applied scoring formula.

The San Francisco rating plan utilizes a simple check form applying uniformly to all similar classes of positions in the various city departments, with the exception of the Police and Fire Departments, for which separate forms have been prepared. The form is filled out by supervisors and double checked by department heads or appointing officers.

The traits included in the form which apply to all employees are as follows: Promptness, Attendance, Quality of Work, Ability to Learn, Co-operation, Dependability, Judgment and Initiative. Additional traits which apply only to certain jobs are: Volume of Work, Contacts with Public, Physical Fitness, Appearance, and Initiative. In the case of supervisory positions, Ability to Train and Organization of Work are included. The five phrases descriptive of the degree of each trait are arranged in the form of a check list, and excellence on the scale is rated sometimes at the beginning and sometimes at the end. Such an arrangement increases the probability that the rater will pay full attention to the descriptive phrases.

On the back of the report is space for the certificate of the reporting officer and of the employee involved. The appointing officer in cases involving probationary employees is required to indicate his impression of the employee's General Fitness to hold the position, and to indicate presence or absence of undesirable characteristics which would make the employee unsuited to the particular job. If the probationary employee is rejected, the appointing officer must indicate specifically why. Ratings are given probationary employees at end of the second and fifth months of service, and give appointing officers a legitimate excuse for rejection of an unsatisfactory applicant by refusal to certify him for permanent tenure.

The probationary period is a definite part of the examination and tests those factors for which no adequate written or oral tests have been devised. Unfortunately, however, appointing officers often fail to recognize their responsibility in checking the probationer's services and too often the probationary period elapses without adequate investigation of the employee's performance on the job. It is hoped that the use of the service rating system for probationers will serve to impress upon the appointing officer his responsibility in determining whether or not the probationary appointee will be a satisfactory permanent employee.

We expect, for permanent employees, to rate once or twice per year.

NEW DEVELOPMENT FOR FIRE MOTOR DRIVER EXAMINATION

WILLIAM E. TRUOG, JR.
Kansas City Personnel Department, Kansas City, Missouri

THE development of a new performance testing procedure for Fire Motor Drivers followed some extensive research by the Personnel Department of the City merit system of Kansas City, Missouri. This examination is based upon a standardized test used by the United States Army Engineers to examine all motor vehicle drivers, and was developed by Amos E Neyhart, Administrative Head, Institute of Public Safety of Pennsylvania State College and Consultant to the American Automobile Association. The Army examination consisted of a short written exercise followed by tests for visual acuity, field of vision, depth perception, and the applicant's reaction to simulated traffic situations The Army performance procedure includes a closed area and a general road test, with an appraisal of the general driving characteristics of the applicant. The driver's performance test developed by the Kansas City Personnel Department is based primarily on the closed area procedure used by the Army Engineers

Five tasks were selected (1) Driving on a straight line—forward and backward, which consisted of lining up the front and rear left wheels on a painted 100-foot line, and driving at normal speed, to keep both front and rear wheels on the line for the entire distance (2) Gauging space when steering in close limits. This is a timed test in which the contestant has to make a 90-degree turn going forward, and then to back up, making the same turn (3) Stopping the car smoothly in 40 feet while going 20 miles an hour. Lines are painted on the street as a guide to begin slowing down and to stop. (4) Stopping the car with the front wheel exactly on a cross painted on

the street This procedure is repeated for the front bumper and also for the rear wheel and bumper. The contestant is penalized doubly for going beyond the line. (5) Parking the car against a curb in a regulation parallel parking

A road test through regular traffic conditions includes an appraisal of the applicant's making a right- and a left-hand turn, of his starting from a standstill while on an upgrade, and finally, of his handling of the car and of his own reactions and emotional status in meeting traffic situations A regular touring car was used for this first test, but there is no doubt that fire department apparatus would provide a much more discriminating result It was found that generally the touring car proved easy for everyone to maneuver through the various operations

One of the best developments of the Neyhart testing procedure is the detailed scoring sheet devised for the use of the examiner Under each task are outlined several items upon which the contestant is to be graded, e.g.

Sample Score Sheet

III Stopping smoothly in 40 feet at 20 miles an hour—Front Bumper
a Driving Forward through Stanchions,
  1 Moves gear shift lever to another position without clashing gears
  2. Keeps an even speed—20 M P H
  3 Moves vehicle continuously—no stops
  4 Steers with certainty—no sudden jerks
  5 Does not stop between stanchions
  6. Does not hit right—left stanchions
  7. Does not race engine.
  8 Does not stall engine.
  9 Stops vehicle smoothly.
  10 Stops front bumper short—over line

A more accurate and discriminating grade is made possible by this system, and the driver may later see the points of the exercise that he failed to perform correctly. The examiner checks those points missed as the operation is being completed and scores each exercise before going to the next It is believed that this has proved a much more objective test than others, and by the use of the detailed scoring sheet, the standards outlining the

course, and lines drawn in the street to guide the contestant, the result depends less on the judgment of the examiner and provides a positive, equal standard for all applicants. It leaves little room for the driver to dispute a point, as there can be no doubt when he hits or knocks down a standard.

The equipment used for this test is an important factor, both for the applicant and the examiner Standards painted white and topped with red flags are used to outline the different courses, supplemented by four-inch white lines painted on the street as a guide to the driver The tasks for this test may all be laid out in an ordinary street

The examination for Fire Motor Driver is divided into four sections, a written examination, performance test, service rating, and rating of experience and training. The written and performance sections are both given a valuation of 30%, with the service rating and rating of experience and training each being valued at 20% of the total. The service rating is also a new development and was used for the first time in this examination. There has been no further opportunity to determine the validity of this new procedure, except for the correlations which have been computed. Following are the correlations between the various sections of the test: Performance and the total score, + .58; performance and written, + .02; written and total score, + .51; performance and service rating, – 17.[*] The low correlation between the performance test scores and written scores is regarded as particularly desirable, since the two sections are testing for different qualities.

The scoring and grading system for the performance test is based on a raw score of 100. The test is divided into ten tasks, each one having a valuation of 10. Points are deducted for each item of the operation that was performed incorrectly. A distribution table, based on the scores, is made up, and a passing point is set. A conversion table is computed on the basis of the 30% valuation set for the performance test The total raw score is then converted and added to the rest of the score for the entire examination.

The Army Engineers and the Fire Department worked in full cooperation with the Personnel Department in developing the examination. With some further refinements, it is believed that this will constitute a highly valid and successful testing procedure

---

[*] Since this is the first use of the service rating on an examination, we have had no measurement of the validity of it. For this reason, the correlation of – 17 should not be regarded as seriously as the other correlations.

MEASUREMENT ABSTRACTS*

Crider, B. "A Battery of Tests for the Dominant Eye." *Journal of General Psychology*, XXXI (1944), 179–190.

In order to meet the need for uniformity of procedures and test materials used to determine eye dominance, the author has developed a battery of seven simple tests for this purpose. Directions for making the test materials, precise instructions for administration, a form for recording performance, and norms based on the records of over 700 subjects are given. The author reports that "coefficients of reliability determined by repeating the tests are all over .98 or .99." Of the population for which norms are given, 53 per cent were definitely right eyed on the battery, 24 per cent were definitely left eyed, 7 per cent impartial-eyed, and the remainder showed left-eyed tendency or right-eyed tendency. *Edith S. Jay*

Griffin, C. H. and Borow, H. "An Engineering and Physical Science Aptitude Test." *Journal of Applied Psychology*, XXVIII (1944), 376–387.

By selecting portions of the *Revised Iowa Physics Aptitude Test*, the *Moore Test of Arithmetic Reasoning*, the *Bennett Test of Mechanical Comprehension*, the *Moore Hall Examination*, and others, an aptitude test was constructed for use in assessing suitability of candidates for training in technical work on the high-school or college level. The six parts of the examination can be administered in 74 minutes of student working time. Percentile norms have been established separately for 6695 men and 2295 women enrolled in war training courses at the Pennsylvania State College. Multiple correlations of the weighted test parts and course achievement were as high as .79. Reliability determined by the split half method, corrected with the Spearman Brown formula, was .96. The subtests are Mathematics, Formulation, Physical Science Comprehension, Arithmetic Reasoning, Verbal Comprehension, and Mechanical Comprehension, representing relatively independent aspects of technical performance. *Edith S. Jay.*

Holzinger, K. J. "The Relationship between the Centroid Method and Spearman's Method." *Journal of Educational Psychology*, XXXV (1944), 347–352.

This article points out the relationship between the centroid method of obtaining factor coefficients and Spearman's 1914 method, as well as the relationship between the centroid method and Spearman's theorem on the correlation of sums. It is shown that if communalities are inserted in the diagonals, Spearman's method for obtaining a factor coefficient is the same as the centroid method. Similarly, the method of correlation of sums, again with the communalities included, is equivalent to the centroid method, it being demonstrated that "centroid coefficients a_j may thus be regarded as correlations between a variable z_j and the total (or average) of the variables $z_1, z_2, \ldots z_n$ projected on the common-factor space." *L. Bouthilet*

Kenfer, Homer. "Simplifying the Scoring Technique of the Bernreuter Personality Inventory." *Journal of Applied Psychology*, XXVIII (1944), 412–413.

Five simplified keys were used to replace the usual scoring technique of the *Bernreuter Personality Inventory*: I. All original values of 3 to −3 were assigned a value of 0, 4 and above became 1, −4 and below became −1. II. Original values of 2 to −2 became 0, 3 and above became 1, −3 and below became −1. III. "?" responses were ignored and all "Yes" and "No" values having a difference of seven (reduced to five on scale B3-I) or more were paired. Positive values became 1 and negative values became −1. IV. "?" responses were ignored and all positive items were counted as 1 if there was a difference of six (reduced to five on scale B3-I) or more between the "Yes" and "No" answers. V. All positive responses were counted as 1. Zero and negatives were ignored. After the new scores were derived, Pearson correlations were worked for each scale with results from each key. Seventeen of the correlations of the five keys with original percentile ranks on the four scales range from .82 to .91. Probable errors range from ±.011 to ±.032.

It is concluded that the best new key for each trait could be used for quick location of extreme cases, the middle fifty per cent considered "normal," and the fifty per cent at the two extremes rescored with the regular Bernreuter keys. The new keys effect a saving of more than 90 per cent in scoring time. *Marion H. Groves*

---

* Edited by Forrest A. Kingsbury

Lundin, R. W. "A Preliminary Report on Some New Tests of Musical Ability." *Journal of Applied Psychology*, XXVIII (1944), 393–396.

This describes a battery of 5 tests devised to measure directly and objectively musical abilities, learned rather than innate. The tests—recorded musical items on which "same" or "different" judgments were requested for interval discrimination, melodic transposition, harmonic transposition, melodic sequences and harmonic sequences—were given to 2 groups of subjects: 60 music students from DePauw and Indiana Universities and 100 students from undergraduate classes in psychology at the latter. The total scores showed a reliability of .73, but the reliabilities of the separate tests were not so high. The only criterion for validation was the author's specially constructed graphic rating scale by which the professors of music graded their students. Finding a significant difference between the 2 groups, the author believes the tests do discriminate between those with and without musical ability. *Vernon S. Tracht*

McCarthy, D. "A Study of the Reliability of the Goodenough Drawing Test of Intelligence." *Journal of Psychology*, XVIII (1944), 201–216.

The Goodenough *Drawing Test of Intelligence* was given two times, a week apart, to 386 third and fourth-grade children. Each test was scored three times, twice by the same scorer, and once by different scorers. Tests scored by the same person gave a correlation of .94, with 12.4 per cent of the cases having a discrepancy of one year or more. Correlation between scores by different scorers was .90, but discrepancies of one year or more were found in 25.3 per cent of the tests. Moreover, although the consistency on two tests taken at different times and scored by one person was .68, there was a discrepancy of at least one year in 41.7 per cent of the cases. The odd even reliability computed with the Spearman Brown prophecy formula was .89. The results, demonstrating both the subjectivity of scoring and the variability in performance over a short interval, show the need for caution in using the scale for individual diagnosis. *L. Bouthilet.*

McClelland, David C. "Simplified Scoring of the Bernreuter Personality Inventory." *Journal of Applied Psychology*, XXVIII (1944), 414–419.

The scoring of the *Bernreuter Personality Inventory* was simplified by assigning a value of 1 to all answers weighted 3 or above and a value of −1 to all answers weighted −3 or below; answers weighted between these limits were ignored. Correlations between full and simplified scoring of the inventories of 114 college men on five traits were computed; they were above .95 for B1N, B4D, F1C, and F2S, and .84 for B2S. Formulas for converting a short score into a full score and a table of short score percentile norms for college men are included. The time required to short-score an inventory is about 1 minute per trait. *Marion H. Groves*

Mellone, M. A. "A Factorial Study of Picture Tests for Young Children." *British Journal of Psychology*, XXXIV (1944), 9–16.

A battery of 14 picture tests, a word reading test, and a mechanical arithmetic test was given to 414 seven-year-old children, 218 boys and 196 girls. The intercorrelations between tests were factored by the centroid method into 3 common factors and specifics. After rotation, the 3 factors were identified as (1) a general factor found in all of the picture tests and slightly in the other tests, (2) a scholastic factor present in the complete battery but not in the picture-tests battery alone, and (3) a space factor, appearing in some of the picture tests, which showed a sex difference, evident only in the boys. *L. Bouthilet*

Odell, C. W. "The Scoring of Continuity or Rearrangement Tests." *Journal of Educational Psychology*, XXXV (1944), 352–356.

Tests in which the task is to rank up to six items in some designated order can be scored quickly and in a theoretically sound manner by using a table presented as an aid in shortening the calculations. The table is divided into sections which enable the scorer to penalize with minus scores for poorer than chance arrangement, to give zero scores to all arrangements negatively correlated with the correct order, or to score by the better, but longer, *p* method which requires that differences between correct ranks and pupil's responses be squared and summed. *Edith S. Jay*

Postwar Counseling on the College Campus. Reflections from the Institute on Student Personnel Work held at the University of California, Los Angeles, Summer Session, 1944. Published by Western Personnel Service, Pasadena, California, 1944.

This is a 20-page pamphlet prepared by a committee at the Institute. It aims to present the different points of view in a series of short articles. E. G. Williamson as "Leader of the Institute" emphasizes the need for a well planned integrated personnel program. The significance of "the contributions of the students themselves to their own development" is discussed by Jessie E. Gibson, "Dean of Women." The "Personnel Administrator," Karl W. Onthank, in reviewing the various papers read at the meetings brings out many of the practical considerations that must be met in counseling after the war. F. T. Perkins, as "Psychologist," comments favorably on the fact that many diverse agencies are interested in the problems of counseling, but suggests that the faculty of colleges should be brought into the personnel program to a greater extent. Finally, Ruth P. McLain, representing the "Layman," was impressed by the evidence that not only will colleges be faced with great demands in the postwar years, but that the new profession of personnel work will be of much value in meeting these demands. *L. Bouthilet.*

Rabin, A. I. "Test Constancy and Variation in the Mentally Ill." *Journal of General Psychology*, XXXI (1944), 231–239.

The *Wechsler-Bellevue Intelligence Scales* were used on 60 New Hampshire State Hospital adult patients to determine the relative test-retest consistency of mental patients as well as the sensitivity of the measuring instrument. The results revealed a correlation of .84 between them and compared favorably with those obtained from different tests on normal persons, tending to disprove the belief that test results of psychotics are very unstable. The coefficient also indicated a high degree of reliability for the whole scale itself, according to the author's conclusions. Slight but consistent increases in total and individual retest scores were observed, especially on the performance part of the scale, the magnitude being directly related to the extent of time between test and retest. *Vernon S. Tracht.*

Tuckman, Jacob. "A Study of the Reliability of the Minnesota Rate of Manipulation Test by the Split-Half and Test-Retest Methods." *Journal of Applied Psychology*, XXVIII (1944), 388–392.

Since the manual for the *Minnesota Rate of Manipulation Test* presents no reliability data the author undertook to determine the reliability of the test. The split half and test-retest techniques were used. Subjects for the split-half study (odd-even of the four test trials) were 386 men (17 to 56, M age 21.9), 319 women (17 to 45, M age 21.8), 143 boys (13.4 to 18.9, M age 15.8), and 111 girls (13.9 to 18.4, M age 15.6). After applying the Spearman Brown prophecy formula the corrected coefficients for both Placing and Turning range from .91 to .97 with no difference among the four groups. S.D.'s range from 19.9 to 24.3. An additional 100 high school students—51 boys and 49 girls—with a mean age of 16.3 were subjects for

both a split-half and a test retest study. The corrected split half coefficients for all possible combinations of single trials range from .82 to .93. The interval between the initial test and the retest varied from 1 to 14 days with a median of 7 days. In the test retest study a progressive though small decrease in time for each successive trial for both Placing and Turning on the initial and final test is revealed. All subjects were faster on the retest for both Placing and Turning. The D/σ difference between total score on initial and final test is 6.7 for Placing and 10.4 for Turning. *Marion H. Groves.*

NEWS NOTES

Regional meetings of the Council of Guidance and Personnel Associations will be held as follows

ATLANTA, GEORGIA

Dates: Feb. 22, 23, 24, 1945
Headquarters: Ansley Hotel
Chairman Dr. W D Perry, Director, Military and Vocational Information, 207 South Building, University of North Carolina, Chapel Hill, N C.
Program theme: A Post-War Picture of the South.

CHICAGO, ILLINOIS

Dates: Feb. 15, 16, 17, 1945
Headquarters, Morrison Hotel
Chairman: Mr E. L. Kerchner, Supervisor, Employment Certificate Office, Board of Education, 228 North LaSalle St., Chicago, Ill.
Program theme: Mobilization of Community Counselling Services.

DENVER, COLORADO

Dates Feb. 26, 27, 1945
Headquarters: Albany Hotel
Chairman: Mr Dwight C. Baird, Counselor, Occupational Information and Guidance Service, State Board for Vocational Education, Room 210 State Office Building, Denver, Colo.
Program theme. Mobilization of Community Counseling Services.

NEW YORK, NEW YORK

Dates: Feb. 21, 22, 1945
Headquarters: Not yet announced.
Chairman: Dr. Forrest H. Kirkpatrick, Director, Personnel Administration, Radio Corporation of America, Camden, N J
Program theme Mobilization of Community Counseling Services.

SAN FRANCISCO, CALIFORNIA

Plans incomplete. Announcements will be made later.

The membership of A.C.P.A. will be circularized about further details in the near future.

Lucile Brown, who is with the American Red Cross, is now assigned to recreational work in rest centers. She writes that one aspect of her job is to plan tours to points of historical and cultural interest in southern and central Italy

---

Karl Cowdery, Associate Registrar, Stanford University, died in the early fall. Dr. Cowdery had been a member of A.C.P.A. for many years and had served as President, as Editor of the Yearbook, and as Chairman of many important committees. His death is a great loss to the organization and to his many friends.

Forrest Kirkpatrick has been appointed consultant on personnel administration for the Department of State.

NATIONAL INSTITUTE OF EDUCATION
LIBRARY & DOCUMENTATION
Unit (NCERT)
No. J-2727
Date 4-3-77

Grace E Manson has accepted an appointment in the Civilian Personnel Research Subsection of the Classification and Replacement Branch of the Adjutant General's Office at 270 Madison Avenue, New York, N Y

Harry W Seamans, formerly General Secretary of the Penn State Christian Association, Pennsylvania State College, and recently Community Participation Adviser, National Housing Agency, Office of the Administrator, Washington, D C, has become Chief of Employee Relations for the O P.A., Washington, D C

Western Personnel Service has published a report on "Post-war Counseling on the College Campus," which gives reflections from the Institute on Student Personnel Work held last summer at the University of California, Los Angeles, in collaboration with Western Personnel Service Contributors are E. G. Williamson, Dean of Students, University of Minnesota, Jessie E. Gibson, President, California Association of Deans of Women, Karl W Onthank, Dean of Personnel Administration, University of Oregon, F T Perkins, Associate Professor of Psychology, Claremont Colleges, and Ruth P McLain, Secretary, Board of Directors, Western Personnel Service.

Mrs Chase Going Woodhouse, Connecticut College for Women, was elected to Congress in the recent elections.

INDEX FOR VOLUME IV

Adams, William Michael
Prediction of Scholastic Success in Colleges of Law; II An Investigation of Pre-Law Grades and Other Indices of Law-School Aptitude 13
Adkins, Dorothy C
Test Construction in Public Personnel Administration 141
Allen, Mildred M.
Relationship Between Kuhlman-Anderson Intelligence Tests in Grade 1 and Academic Achievement in Grades 3 and 4 161
Andrews, Elizabeth Gordon
Guidance-Survey of Student Problems 209
Beers, Fred S (with Cecil R Brolyer)
Major Strategy versus Minor Tactics in Merit Administration 97
Berdie, Ralph F
Judgments in Counseling 35
Bolanovich, Daniel J (with Charles H. Goodman)
A Study of the Kuder Preference Record 315 ✓
Brolyer, Cecil R (with Fred S Beers)
Major Strategy versus Minor Tactics in Merit Administration 97
Burnham, Paul S. (with Albert B. Crawford)
Trial at Yale University of the Armed Forces Institute General Educational Development Tests 261
Carlin, Albert V. (with Frederic M. Lord)
A Report on Scholarship Examinations Given in Latin American Countries for the Selection of Students to be Trained in Meteorology 69
Crawford, Albert B. (with Paul S. Burnham)
Trial at Yale University of the Armed Forces Institute General Educational Development Tests 261
Dunkel, Harold B.
An Inventory of Students' General Goals in Life 87
Englehart, Max D
How Teachers Can Improve Their Tests 109
Ewing, Alfred M
The Use of Standardized Reading Tests in Teachers Colleges 225
[illegible]
[illegible] 21
[illegible]
[illegible] 179

Goodman, Charles H.
Prediction of College Success by Means of Thurstone's Primary Abilities Tests 125
Goodman, Charles H. (with Daniel J Bolanovich)
A Study of the Kuder Preference Record 315
Harrison, L D
A Validating Study of the Oberlin Vocational Interest Inquiry 199
Hewitt, Forrest E.
Guidance through Self-Appraisal 245
Lehman, Ruth T
Interpretation of the Kuder Preference Record for College Students of Home Economics 217
Link, Henry C.
The Definition of Social Effectiveness and Leadership through Measurement 57
Lord, Frederic M (with Albert V. Carlin)
A Report on Scholarship Examinations Given in Latin American Countries for the Selection of Students to be Trained in Meteorology 69
McDaniel, J. W (with Wm. A Reynolds)
A Study of the Use of Mechanical Aptitude Tests in the Selection of Trainees for Mechanical Occupations 191
Manson, G. E (with G L. Freeman)
A Technique for Evaluating Assembled Evidence of Potential Leadership Ability 21
Peterson, Shailer
The Word-Dexterity Test, A Better Measure of College Aptitude 307
Reynolds, Wm. A. (with J. W McDaniel)
A Study of the Use of Mechanical Aptitude Tests in the Selection of Trainees for Mechanical Occupations 191
Smith, Ireneus S.
Developing a Service Rating System 327
Thompson, Claude Edward
Personality and Interest Factors in Dental School Success 299
Todd, J E
Measurement in the Continuous Selection and Counseling of Students in a College of Physical Education and Social Work 233
Toops, Herbert A
The Criterion 271
Truog, William E., Jr
New Development for Fine Motor Driver Examination 339
Williamson, E G
Student Activities—An Integral Part of Student Personnel Work 3